526 Rec'd PCT/PTO 16 AUG 2000

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER 48792

U.S. APPLICATION NO. (If known, see 37 CFR 1.5) 09/622419

| INTERNATIONAL APPLICATION NO. | INTERNATIONAL FILING DATE | PRIORITY DATE CLAIMED |
|---|---|---|
| PCT/EP 99/01052 | 17 February 1999 | 19 February 1998 |

TITLE OF INVENTION: PROCESS FOR PREPARING BIOTIN

APPLICANT(S) FOR DO/EO/US Hartwig SCHROEDER

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. /X/ This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.

2. / / This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.

3. /X/ This express request to begin national examination procedures (35 U.S.C.371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).

4. /x / A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.

5. /X/ A copy of the International Application as filed (35 U.S.C. 371(c)(2)).

   a./X/ is transmitted herewith (required only if not transmitted by the International Bureau).
   b./ / has been transmitted by the International Bureau.
   c./ / is not required, as the application was filed in the United States Receiving Office (RO/USO).

6. /X/ A translation of the International Application into English (35 U.S.C. 371(c)(2)).

7. / / Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)).

   a./ / are transmitted herewith (required only if not transmitted by the International Bureau).
   b./ / have been transmitted by the International Bureau.
   c./ / have not been made; however, the time limit for making such amendments has NOT expired.
   d./ / have not been made and will not be made.

8. / / A translation of the amendments to the claims under PCT Article 19(35 U.S.C. 371(c)(3)).

9. /X/ An oath or declaration of the inventor(s)(35 U.S.C. 171(c)(4)).

10./ / A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11./ / An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12./X/ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.

13./X/ A FIRST preliminary amendment.
   / / A SECOND or SUBSEQUENT preliminary amendment.

14./ / A substitute specification.

15./ / A change of power of attorney and/or address letter.

16./x / Other items or information.
International Search Report
International Preliminary Examination Report

534 Rec'd PCT/PTO 16 AUG 2000

09/622419

U.S. Appln. No. (If Known) | INTERNATIONAL APPLN. NO. PCT/EP 99/01052 | ATTORNEY'S DOCKET NO. 48792

| | CALCULATIONS | PTO USE ONLY |
|---|---|---|
| 17. /X/ The following fees are submitted | | |
| BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5)): Search Report has been prepared by the EPO or JPO......................................$840.00 | 840.00 | |
| International preliminary examination fee paid to USPTO (37 CFR 1.482)......................................$750.00 | | |
| No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2))........................$700.00 | | |
| Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO .................$ 970.00 | | |
| International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied pro -visions of PCT Article 33(2)-(4)....................$96.00 | | |
| ENTER APPROPRIATE BASIC FEE AMOUNT = $ | 840.00 | |
| Surcharge of $130.00 for furnishing the oath or declaration later than / / 20 / /30 months from the earliest claimed priority date (37 CFR 1.492(e)). | | |

| Claims | Number Filed | Number Extra | Rate | | |
|---|---|---|---|---|---|
| Total Claims | 14 -20 | | X$18. | | |
| Indep.Claims | 3 -3 | | X$78. | | |
| Multiple dependent claim(s)(if applicable) | | | +270. | | |
| TOTAL OF ABOVE CALCULATION | | | = | 840.00 | |
| Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28). | | | | | |
| SUBTOTAL | | | = | 840.00 | |
| Processing fee of $130. for furnishing the English translation later than / /20 / /30 months from the earliest claimed priority date (37 CFR 1.492(f)). + | | | | | |
| TOTAL NATIONAL FEE | | | = | 840.00 | |
| Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) $40.00 per property = | | | | | |
| TOTAL FEES ENCLOSED | | | = $ | 840.00 | |

Amount to be refunded: $
Charged $

a./X/ A check in the amount of $ 840. to cover the above fees is enclosed.

b./ / Please charge my Deposit Account No.____________in the amount of $__________ to cover the above fees. A duplicate copy of this sheet is enclosed.

c./X/ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 11-0345. A duplicate copy of this sheetis enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b) must be filed and granted to restore the application to pending status.

Herbert B. Keil
SIGNATURE

SEND ALL CORRESPONDENCE TO:
KEIL & WEINKAUF
1101 Connecticut Ave., N.W.
Washington, D. C. 20036

Herbert B. Keil
NAME
18,967
Registration No.

09/622419

534 Rec'd PCT/PTO 16 AUG 2000

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of )
SCHROEDER et al. ) BOX PCT
)
International Application )
PCT/EP 99/01052 )
)
Filed: February 17, 1999 )
)
For: PROCESS FOR PREPARING BIOTIN

PRELIMINARY AMENDMENT

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

Sir:

Prior to examination, kindly amend the above-identified application as follows:

IN THE CLAIMS

Claim 3, line 1, delete "or 2".

Claim 4, line 1, delete "any of claims 1 to 3" and insert --claim 1--.

Claim 5, line 1, delete "any of claims 1 to 4" and insert --claim 1--.

Claim 6, line 1, delete "any of claims 1 to 5" and insert --claim 1--.

Claim 9, line 1, delete "or 8".

Claim 10, line 1, delete "any of claims 7 to 9" and insert --claim 7--.

Claim 11, line 2, delete "any of claims 7 to 10" and insert --claim 7--.

Claim 14, lines 1 and 2, delete "any of claims 7 to 10" and insert --claim 7--.

REMARKS

The claims have been amended to eliminate multiple dependency and to put them in better form for U.S. filing. No new matter is included.

Favorable action is solicited.

Respectfully submitted,

KEIL & WEINKAUF

HBKeil

Herbert B. Keil
Reg. No. 18,967

KEIL & WEINKAUF
1101 CONNECTICUT AVENUE, N.W.
WASHINGTON, D.C. 20036

0050/48792 2/PRTS 09/622419
534 Rec'd PCT/PTO 16 AUG 2000

# Process for preparing biotin

The invention relates to a gene construct which contains an S-adenosylmethionine synthase gene, having the sequence SEQ ID No. 1, and a biotin biosynthesis gene bioS1, bioS2 and/or bioS3, having the sequences SEQ ID No.3, SEQ ID No.5 and SEQ ID No.7, respectively, and, where appropriate, at least one further biotin synthesis gene sequence selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR. The invention furthermore relates to organisms which contain this gene construct and to the use of the gene construct for preparing biotin, and also to a process for preparing biotin.

As a coenzyme, biotin (Vitamin H) plays an essential role in enzyme-catalyzed carboxylation and decarboxylation reactions. Biotin is consequently an essential factor in living cells. Almost all animals and some microorganisms have to take up biotin from the exterior since they are unable to synthesize biotin themselves. Biotin is therefore an essential vitamin for these organisms. By contrast, bacteria, yeasts and plants are able themselves to synthesize biotin from precursors (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326, DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5).

The synthesis of biotin has been investigated in bacterial organisms, especially in the Gram-negative bacterium Escherichia coli and in the Gram-positive bacterium Bacillus sphaericus (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326). Pimelyl-CoA (PmCoA), which is derived from fatty acid synthesis, has previously been regarded as the first known intermediate in E. coli (DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5 1996). Up to now, the route by which this biotin precursor is synthesized in E. coli has to a large extent been unknown (Lemoine et al., Mol. Microbiol. 19, 1996: 645 - 647). bioC and bioH have been identified as being two genes whose corresponding proteins are responsible for the synthesis of Pm-CoA. The enzymic functions of the gene products, i.e. BioH and BioC, have hitherto been unknown (Lemoine et al., Mol. Microbiol. 19, 1996: 645 - 647, DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5). Pm-CoA is converted into biotin in four further enzymic steps. BioF first of all condenses the Pm-CoA with alanine to form 7-keto-8-aminopelargonic acid (KAPA). The KAPA is then converted into 7,8-diaminopelargonic acid (DAPA)

by BioA. Following an ATP-dependent carboxylation reaction, the next step leads to dethiobiotin (DTB) and is catalyzed by BioD. The DTB is converted into biotin in the last step. This step is catalyzed by BioB. The chemical and enzymic mechanisms involved in the conversion of DTB into biotin are so far only incompletely understood and clarified.

The conversion of DTB into biotin has so far only been characterized in bacterial and plant cell extracts (WO94/8023, EP-B-0 449 724, Sanyal et al. Arch. Biochem. Biophys., Vol. 326, No. 1, 1996: 48 - 56 and Biochemistry 33, 1994: 3625 - 3631, Baldet et al. Europ. J. Biochem. 217, 1, 1993: 479 - 485, Méjean et al. Biochem. Biophys. Res. Commun., Vol. 217, No. 3, 1995: 1231 - 1237, Ohshiro et al., Biosci. Biotechnol. Biochem., 58, 9, 1994: 1738 - 1741).

In vitro studies have demonstrated that low molecular weight factors such as NADPH, cysteine, thiamine, $Fe^{2+}$, asparagine, serine, fructose 1-6-bisphosphate and S-adenosylmethionine are able to stimulate the synthesis of biotin (Ohshiro et al., Biosci. Biotechnol. Biochem., 58, 9, 1994: 1738 - 1741, Birch et al., J. Biol. Chem. 270, 32, 1995: 19158 - 19165, Ifuku et al., Biosci. Biotechnol. Biochem., 59, 2, 1995: 185 - 189, Sanyal et al. Arch. Biochem. Biophys. 326, 1, 1996: 48 - 56).

In addition to these low molecular weight factors, other proteins have been identified which stimulate the synthesis of biotin from DTB in the presence of BioB. These proteins are flavodoxin and flavodoxin NADPH reductase (Birch et al., J. Biol. Chem. 270, 32, 1995: 19158 - 19165, Ifuk et al., Biosci. Biotechnol. Biochem., 59, 2, 1995: 185 - 189, Sanyal et al., Arch. Biochem. Biophys. 326, 1, 1996: 48 - 56). Other proteins which stimulate biotin synthesis are the genes bioS1 and bioS2, which are described in the German application having the application number 197.31274.8 (Priority 22.7.97).

Differing results have been obtained with regard to the origin of the sulfur in the biotin molecule. Investigations into the synthesis of biotin in whole cell extracts showed that radioactivity was incorporated into biotin in the presence of $^{35}S$-labeled cysteine; it was not possible to demonstrate incorporation of sulfur into the biotin molecule when either $^{35}S$-labeled methionine or S-adenosylmethionine was used (Ifuku et al., Biosci. Biotechnol. Biochem. 59, 2, 1995: 184 - 189, Birch et al., J. Biol. Chem.270, 32, 1995: 19158 - 19165).

The genes which encode the described proteins, i.e. bioF, bioA, bioD, and bioB, are encoded in E. coli on a bidirectional operon. This operon is located between the λ attachment site and the uvrB gene locus at approx. 17 minutes on the E. coli chromosome (Berlyn et al. 1996: 1715 - 1902). A further two genes, one of which, i.e. bioC, already possesses described functions in the synthesis of Pm-CoA, are additionally encoded on this operon, whereas it has not so far been possible to assign any function to an open reading frame which is located downstream of bioA (WO94/8023, Otsuka et al., J. Biol. Chem. 263, 1988: 19577 - 85). Highly conserved homologues to the E. coli proteins BioF, BioA, BioD and BioB have been found in B. sphaericus, B. subtilis, Syneccocystis sp. (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326, Bower et al., J. Bacteriol. 175, 1996: 4122 - 4130, Kaneko et al., DNA Res. 3, 3, 1996: 109 - 136), archaeobacteria such as Methanococcus janaschi, and yeasts such as Saccharomyces cerevisiae (Zhang et al., Arch. Biochem. Biophys. 309, 1, 1994: 29 - 35) or in plants such as Arabidopsis thaliana (Baldet et al., C. R. Acad. Sci. III, Sci. Vie. 319, 2, 1996: 99 - 106).

In the two Gram-positive microorganisms which have so far been investigated, the synthesis of Pm-CoA appears to proceed in a different manner from that in E. coli. It was not possible to find any homologues of bioH and bioC (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326).

Biotin is an optically active substance which has three centers of chirality. It has hitherto only been prepared economically by way of an expensive, multi-step chemical synthesis.

As an alternative to this chemical synthesis, a large number of attempts have been made to construct a fermentative process for preparing biotin using microorganisms. Cloning the biotin operon onto multi-copy-plasmids has been successfully used to increase biotin synthesis in microorganisms which have been transformed with these genes. A further increase in biotin synthesis was achieved by deregulating biotin gene expression by means of selecting birA mutants (Pai C. H., J. Bacteriol. 112, 1972: 1280 - 1287). Combination of the two approaches, that is expressing the plasmid-encoded biosynthesis genes in a regulation-deficient strain (EP-B-0 236 429), increased productivity still further. In this context, the biotin operon can either remain under the control of its native bidirectional promoter (EP-B-0 236 429) or else its genes can be brought under the control of a promoter which can be regulated externally (WO94/08023).

The approaches which have so far been pursued for producing biotin fermentatively in E. coli have not achieved any economically adequate productivity.

It is an object of the present invention to develop an industrial fermentative process for producing biotin which exhibits as high a biotin synthesis as possible.

We have found that this object is achieved by the process according to the invention for producing biotin, in which process an S-adenosylmethionine synthase (SAM synthase) gene, having the sequence SEQ ID No. 1, and at least one further biotin biosynthesis gene bioS1, bioS2 or bioS3, having the sequences SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No.7, and also their functional variants, analogues or derivatives, are expressed in a prokaryotic or eukaryotic host organism which is able to synthesize biotin, this organism is cultured and the synthesized biotin is used directly after separating off the biomass or after purifying the biotin.

The genes used in the process according to the invention, i.e. the SAM synthase gene having the sequence SEQ ID No. 1 and the biotin biosynthesis genes bioS1, bioS2 and bioS3 having the sequences SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No.7 are kept in the SwissProt-data base under accession numbers P04384 (metK), U29581 (bioS1), P39171 (bioS2) and D90811 (bioS3). A number of homologues to E. coli MetK are described in the data base. These homologues include organisms such as other eubacteria (e.g. H. influenzae, and B. subtilis), and also eukaryotes (e.g. yeasts: S. cerevisiae, Planta: P. deltoides, Arthropoda: D. melanogaster, and Mammalia: R. norvegicus).

The productivity of the biotin biosynthesis can be increased markedly by expressing one or more of the SAM synthase gene, having the sequence SEQ ID No. 1, and its functional variants, analogues or derivatives in combination with at least one of the biotin synthesis genes bioS1, bioS2 or bioS3, having the sequences SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No.7, and also their functional variants, analogues or derivatives, in a prokaryotic or eukaryotic host organism. A combination of the SAM synthase gene and bios1 is preferably used for the expression. At least one further biotin gene selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY and bioR is advantageously expressed at the same time in order to increase the biotin synthesis still further. Expression of the genes increases the synthesis of biotin by at least a factor of 2 as compared with the control without these genes, preferably by a

factor which is greater than 3.

The genes used in the process according to the invention, i.e. the SAM synthase gene having the nucleotide sequence SEQ ID No. 1, the bioS1 gene having the nucleotide sequence SEQ ID No. 3, the bioS2 gene having the nucleotide sequence SEQ ID No. 5 and the bioS3 gene having the nucleotide sequence SEQ ID No.7, which sequences encode the amino acid sequences given in SEQ ID NO: 2, SEQ ID No. 4, SEQ ID No. 6 and SEQ ID No.8, respectively, or their allelic variations, can be obtained following isolation and sequencing. Variants are to be understood as being SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and SEQ ID No.7 variants, respectively, which exhibit from 30 to 100% homology at the amino acid level, preferably from 50 to 100% homology, very particularly preferably from 80 to 100% homology. Allelic variants comprise, in particular, functional variants which can be obtained by the deletion, insertion or substitution of nucleotides from the sequences depicted in SEQ ID NO: 1, SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No.7, with, however, the enzymic activity being retained.

In addition, variants are also to be understood as being functional equivalents of the genes, such as O-acetylserine sulfohydrolase A, O-acetylserine sulfohydrolase B, β-cystathionase (see Flint et al., J. Biol. Chem., Vol. 271, 1996: 16053 - 16067) or nifS and its prokaryotic and eukaryotic homologues, for example from Klebsiella, Candida, yeasts or Caenorhabditis, which are able to assume the enzymic activity of bioS1, bioS2 or bioS3 in the synthesis of biotin.

Functional analogues of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 and SEQ ID No. 7 are to be understood as being, for example, their prokaryotic or eukaryotic homologues, such as bacterial, fungal, plant, animal or human homologues. In addition, analogues are also to be understood as being truncated sequences, or single-stranded DNA or RNA from coding and non-coding DNA sequences.

Derivatives are to be understood, for example, as being promoter variants. The promoters, which are placed upstream of the given nucleotide sequences, can be altered by means of one or more nucleotide substitutions, or by means of (an) insertion(s) and/or deletion(s) without, however, the functionality or activity of the promoters being impaired. In addition, the activities of the promoters can be increased by means of altering their sequences, or the promoters can be completely replaced by more active

promoters, including those from organisms of a different species.

Derivatives are also to be understood as being variants whose nucleotide sequences have been altered in the region from -1 to -30 upstream of the start codon such that expression of the gene and/or expression of a protein is increased. This is advantageously effected by altering the Shine-Dalgarno sequence.

All Gram-negative or Gram-positive bacteria which synthesize biotin are, in principal, suitable for use as prokaryotic host organisms in the process according to the invention. Gram-negative bacteria which may be mentioned by way of example are Enterobacteriaceae such as the genera Escherichia, Aerobacter, Enterobacter, Citrobacter, Shigella, Klebsiella, Serratia, Erwinia or Salmonella, Pseudomonadaceae such as the genera Pseudomonas, Xanthomonas, Burkholderia, Gluconobacter, Nitrosomonas, Nitrobacter, Methanomonas, Comamonas, Cellulomonas or Acetobacter, Azotobacteraceae such as the genera Azotobacter, Azomonas, Beijerinckia or Derxia, Neisseriaceae such as the genera Moraxella, Acinetobacter, Kingella, Neisseria or Branhamella, the Rhizobiaceae such as the genera Rhizobium or Agrobacterium, or the Gram-negative genera Zymomonas, Chromobacterium or Flavobacterium. Gram-positive bacteria which may be mentioned by way of example are the endospore-forming Gram-positive aerobic or anaerobic bacteria such as the genera Bacillus, Sporolactobacillus or Clostridium, the coryneform bacteria such as the genera Arthrobacter, Cellulomonas, Curtobacterium, Corynebacterium, Brevibacterium, Microbacterium or Kurthia, the Actinomycetales such as the genera Mycobacterium, Rhodococcus, Streptomyces or Nocardia, the Lactobacillaceae such as the genera Lactobacillus or Lactococcus, or the Gram-positive cocci such as the genera Micrococcus or Staphylococcus.

Preference is given to using bacteria of the genera Escherichia, Citrobacter, Serratia, Klebsiella, Salmonella, Pseudomonas, Comamonas, Acinetobacter, Azotobacter, Chromobacterium, Bacillus, Clostridium, Arthrobacter, Corynebacterium, Brevibacterium, Lactococcus, Lactobacillus, Streptomyces, Rhizobium, Agrobacterium or Staphylococcus in the process according to the invention. Particular preference is given to genera and species such as Escherichia coli, Citrobacter freundii, Serratia marcescens, Salmonella typhimurium, Pseudomonas mendocina, Pseudomonas aeruginosa, Pseudomonas mutabilis, Pseudomonas chlororaphis, Pseudomonas fluorescens, Comamonas acidovorans, Comamonas testosteroni, Acinetobacter calcoaceticus, Azotobacter vinelandii, Chromobacterium violaceum, Bacillus subtilis, Bacillus sphaericus, Bacillus stearothermophilus, Bacillus

pumilus, Bacillus licheniformis, Bacillus amyloliquefaciens, Bacillus megaterium, Bacillus cereus, Bacillus thuringiensis, Arthrobacter citreus, Arthrobacter paraffineus, Corynebacterium glutamicum, Corynebacterium primorioxydans, Corynebacterium sp., Brevibacterium ketoglutamicum, Brevibacterium linens, Brevibacterium sp., Streptomyces lividans, Rhizobium leguminosarum or Agrobacterium tumefaciens. Advantageously, use is made of bacteria which already exhibit an elevated natural production of biotin.

The taxonomic position of the listed genera has been subject to considerable change in recent years and is still in a state of flux as false genera and species names are corrected. Because of these taxonomic regroupings, which have been frequently required in the past, of the said genera within bacterial systematics, families, genera and species other than those mentioned above are also suitable for the process according to the invention.

All biotin-synthesizing organisms, such as fungi, yeasts, plants or plant cells, are, in principal, suitable for use as eukaryotic host organisms in the process according to the invention. Yeasts which may preferably be mentioned are the genera Rhodotorula, Yarrowia, Sporobolomyces, Saccharomyces or Schizosaccharomyces. Particular preference is given to the genera and species Rhodotorula rubra, Rhodotorula glutinis, Rhodotorula graminis, Yarrowia lipolytica, Sporobolomyces salmonicolor, Sporobolomyces shibatanus or Saccharomyces cerevisiae.

In principal, all plants can be used as the host organism, with preference being given to plants which play a role in animal nutrition or human nutrition, such as corn, wheat, barley, rye, potatoes, peas, beans, sunflowers, palms, millet, sesame, copra or rape. Plants such as Arabidopsis thaliana or Lavendula vera are also suitable. Particular preference is given to plant cell cultures, plant protoplasts or callus cultures.

Microorganisms such as bacteria, fungi, yeasts or plant cells which are able to secrete biotin into the growth medium, and which, where appropriate, already additionally exhibit an increased natural synthesis of biotin, are advantageously used in the process according to the invention. Advantageously, these organisms can also be defective with regard to the regulation of their biotin biosynthesis; i.e. this synthesis is either not regulated or only regulated to a very reduced extent. This regulatory defect results in these organisms already possessing a substantially increased biotin productivity. Such a regulatory defect is known, for example, from Escherichia coli in the form

of birA-defect mutants and should preferably be present in the cells as a defect which can be induced by external influences, for example as a defect which is temperature-inducible. In principal, organisms which do not exhibit any natural biotin production can also be used, once they have been transformed with the biotin genes.

In order to increase biotin productivity as a whole still further, the organisms in the process according to the invention should advantageously also harbor at least one further biotin gene selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR. Advantageously, those genes which stimulate biotin synthesis can also be present in the cell in combination with the sequences SEQ ID No. 1 , SEQ ID No. 3, SEQ ID No.5 or SEQ ID No.7 and their combinations. Examples of genes which stimulate biotin synthesis are the flavoredoxin gene and the flavoredoxin reductase gene. This additional gene, or these additional genes, can be present in the cell in one or more copies, like the genes having the sequences SEQ ID No. 1 , SEQ ID No.3, SEQ ID No.5 or SEQ ID No.7 or their combinations. They can be located on the same vector as the sequences SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No.7, or on separate vectors, or else integrated chromosomally. The sequences SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No.7 can also be together on a vector or on separate vectors or be inserted into the genome.

The gene construct according to the invention is to be understood as being the gene sequences of the SAM synthase gene SEQ ID No. 1 and of the biotin synthesis genes SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No.7, and also their functional variants, analogues or derivatives, which were linked functionally to one or more regulatory signals for the purpose of increasing expression of the genes. In addition to these new regulatory sequences, the natural regulation of these sequences can still be present upstream of the actual structural genes and, where appropriate, can have been genetically altered such that the natural regulation has been switched off and expression of genes has been increased. However, the gene construct can also be assembled in a simpler manner, i.e. no additional regulatory signals are inserted upstream of the sequences SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 and/or SEQ ID No.7 and the natural promoter, with its regulation, is not removed. Instead, the natural regulatory sequence is mutated such that regulation by biotin no longer takes place and gene expression is increased. The sequences SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No.7 can be under the regulation of one promoter or under the regulation of separate promoters. Additional, advantageous regulatory elements can also be inserted at the 3' end of the DNA sequences. The

genes having the sequences SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 or SEQ ID No. 7 can be present in the gene construct in one or more copies.

Advantageous regulatory sequences for the process according to the invention are present, for example, in promoters such as the cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, $lacI^{q}$-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, $\lambda$-$P_R$- or $\lambda$-$P_L$-promoters, which are advantageously used in Gram-negative bacteria. Further advantageous regulatory sequences are present, for example, in the Gram-positive promoters amy and SPO2, in the yeast promoters ADC1, MFα , AC, P-60, CYC1, GAPDH or in the plant promoters CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, or nos, or in the ubiquitin promoter or the phaseolin promoter.

In principal, all natural promoters, together with their regulatory sequences, can be used, like the abovementioned promoters, for the process according to the invention. In addition, synthetic promoters can also advantageously be used.

Other biotin genes selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR, which genes can have their own promoter or else can be under the regulation of the promoter of one of the sequences, or under the regulation of the promoter of all the sequences, SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 or SEQ ID No.7, can be present in the gene construct in one or more copies.

For expression in the abovementioned host organism, the gene construct is advantageously inserted into a host-specific vector which makes it possible to achieve optimum expression of the genes in the host. Vectors are well known to the skilled person and can be identified, for example, from the book Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018). In addition to plasmids, the vectors are also to be understood as being all other vectors known to the skilled person, such as phages, viruses, transposons, IS elements, phasmids, cosmids or linear or circular DNA. These vectors can be replicated autonomously in the host organism or replicated chromosomally.

Expression systems are to be understood as being the combination of the host organisms which are mentioned above by way of example and the vectors which are appropriate for the organisms, such as plasmids, viruses or phages, for example plasmids containing the RNA polymerase/promoter system, phages $\lambda$, or Mu or other temperate phages or transposons and/or further advantageous regulatory

sequences.

The term expression systems is preferably to be understood as being the combination of Escherichia coli and its plasmids and phages and the affiliated promoters, and also Bacillus and its plasmids and promoters.

Further 3' and/or 5'-terminal regulatory sequences are also suitable for advantageously expressing SEQ ID No.1, SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No. 7 in accordance with the invention.

These regulatory sequences are intended to make it possible to achieve specific expression of the biotin genes and expression of the protein. Depending on the host organism, this can, for example, mean that the gene is only expressed or overexpressed after induction or that it is expressed and/or overexpressed immediately.

In this context, the regulatory sequences or factors can preferably influence biotin gene expression positively and thereby increase it. For example, the regulatory elements can advantageously be reinforced at the transcriptional level by means of using strong transcription signals such as promoters and/or enhancers. In addition, however, it is also possible to reinforce translation by, for example, improving the stability of the mRNA.

Enhancers are to be understood as being, for example, DNA sequences which bring about increased biotin gene expression by means of improving the interaction between the RNA polymerase and the DNA.

An increase in the proteins (see SEQ ID No.2, SEQ ID No.4, SEQ ID No.6 and SEQ ID No.8) which are derived from the sequences SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and SEQ ID No.7, and in their enzyme activity, as compared with the starting enzymes, can be achieved, for example, by altering the corresponding gene sequences, or the sequences of their homologues, by means of classical mutagenesis, such as UV irradiation, or by treating with chemical mutagens and/or by means of specific mutagenesis such as site-directed mutagenesis, deletion(s), insertion(s) and/or substitution(s). An increased enzyme activity, apart from the described gene amplification, can also be achieved by eliminating factors which repress enzyme biosynthesis and/or by synthesizing active enzymes instead of inactive enzymes.

The process according to the invention advantageously increases the conversion of DTB into biotin, and consequently overall biotin productivity, by means of using the biotin genes having the sequences SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 and SEQ ID No.7, and the combination of the genes having the sequences SEQ ID No.1 and SEQ ID No.5 or SEQ ID No.1 and SEQ ID No.7, preferably the combination of the genes having the sequences SEQ ID No.1 and SEQ ID No.3, which genes are introduced into the organisms by way of their vectors and/or by means of chromosomal cloning.

In the process according to the invention, the microorganisms harboring SEQ ID No.1, SEQ ID No.3, SEQ ID No.5 and/or SEQ ID No.7 are propagated in a medium which enables these organisms to grow. This medium can be a synthetic medium or a natural medium. Use is made of media which are known to the skilled person and which are appropriate for the organism. In order to permit growth of the microorganisms, the media employed contain a carbon source, a nitrogen source, inorganic salts and, where appropriate, small quantities of vitamins and trace elements.

Examples of advantageous carbon sources are sugars, such as monosaccharides, disaccharides or polysaccharides, such as glucose, fructose, mannose, xylose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose, complex sugar sources such as molasses, sugar phosphates, such as fructose-1,6-bisphosphate, sugar alcohols, such as mannitol, polyols, such as glycerol, alcohols, such as methanol or ethanol, carboxylic acids, such as citric acid, lactic acid of acetic acid, fats, such as soy-bean oil or rape-seed oil, or amino acids, such as glutamic acid or aspartic acid, or amino sugars, which can simultaneously be used as a nitrogen source.

Advantageous nitrogen sources are organic or inorganic nitrogen compounds or materials which contain these compounds. Examples are ammonium salts, such as $NH_4Cl$ or $(NH_4)_2SO_4$, nitrates or urea, or complex nitrogen sources such as corn steep liquor, brewer's yeast autolysate, soy-bean flour, wheat gluten, yeast extract, meat extract, casein hydrolysate or yeast or potato protein, which can frequently also be used simultaneously as a nitrogen source.

Examples of inorganic salts are the salts of calcium, magnesium, sodium, manganese, potassium, zinc, copper and iron. Anions of these salts which are to be mentioned in particular are the chloride, sulfate and phosphate ions. An important factor for

increasing productivity in the process according to the invention is the addition of $Fe^{2+}$ or $Fe^{3+}$ salts and/or potassium salts to the production medium.

Where appropriate, further growth factors, such as vitamins or growth promoters, such as riboflavin, thiamine, folic acid, nicotinic acid, pantothenate or pyridoxine, amino acids, such as alanine, cysteine, asparagine, aspartic acid, glutamine, serine, methionine or lysine, carboxylic acids, such as citric acid, formic acid, pimelic acid or lactic acid, or substances such as dithiothreitol, are added to the nutrient medium.

Antibiotics can, where appropriate, be added to the medium in order to stabilize the biotin gene-containing vectors in the cells.

The ratios in which the said nutrients are mixed depends on the nature of the fermentation and is laid down in each individual case. The medium components can all be initially introduced at the beginning of fermentation, after they have been, if necessary, sterilized separately or sterilized together, or else be added subsequently, as required, during fermentation.

The culture conditions are so arranged that the organisms grow optimally and that the best possible yields are achieved. Preferred culture temperatures are from 15 °C to 40 °C. Temperatures of between 25 °C and 37 °C are particularly advantageous. The pH is preferably kept in a range of from 3 to 9. pH values of between 5 and 8 are particularly advantageous. In general, a period of incubation of from 8 to 240 hours, preferably of from 8 to 120 hours, is sufficient. Within this time, the maximum quantity of biotin accumulates in the medium and/or is available after the cells have been disrupted.

The process according to the invention for producing biotin can be carried out continuously or batch-wise or fed-batch-wise. If whole plants are regenerated from the plant cells which have been transformed with the biotin genes, they can, according to the process according to the invention, be grown and propagated perfectly normally.

Examples:

1. Cloning of the S-adenosylmethionine synthase gene (SEQ ID No.1):

Starting from genomic E. coli DNA, the gene which encodes SAM synthase (metK) was amplified from the E. coli chromosome by means of a polymerase chain reaction using two specific oligonucleotides. The DNA which had been amplified in this way was purified, digested with the restriction enzyme Acc65I and inserted into a vector which had been cut with the same enzyme and which enables the gene to be overexpressed in E. coli strains. One of the two oligonucleotides was used to provide the gene construct with optimized translation signals.

a.) Generation of oligonucleotides for amplifying the metK gene from the E. coli chromosome:

metK was to be amplified as an expression cassette which was composed of a ribosome binding site, the start codon of the coding sequence and the stop codon between two restriction enzyme recognition sites. The Acc65I recognition sequence was chosen for both the restriction sites. The metK gene was amplified and cloned using the nucleotides PmetK1 (5'-GCGGTACCAGGTGATATTAAATATGGCAAAAC-3') and PmetK2 (5'-CGGGTACCGATTACTTCAGACCGGCAGC-3').

b.) Implementation of the PCR:

Conditions:

0.5 µg chromosomal DNA from E. coli W3110 was used as a template. The oligonucleotides PmetK1 and PmetK2 were employed at a concentration of in each case 15 pMol. The concentration of the dNTPs was 200 µM. 2.5 U of Pwo DNA polymerase (Boehringer Mannheim) in the manufacturer's reaction buffer were employed as the polymerase. The PCR reaction volume was 100 µl.

Amplifications:

The DNA was denatured at 94 °C for 2 min. The oligonucleotides were then annealed at 55 °C for 30 seconds. The elongation took place at 72 °C for 75 seconds. The PCR reaction was carried out over 30 cycles.

The resulting DNA product, which had a size of approximately 1145 bp, was purified and digested with Acc65I in a suitable buffer.

c.) Cloning of metK in an expression vector

2 µg of the vector pHS1 (construction was described in DE 197.31274.8, priority 22.7.97, Example 1, pages 14 to 17) were digested with Acc65I and dephosphorylated using shrimp alkaline phosphatase (SAP) (Boehringer Mannheim). After the SAP had been denatured, vector and fragment were ligated, in a molar ratio of 1:3, using the Rapid DNA Ligation kit in accordance with the manufacturer's instructions. The ligation mixture was transformed into strain E. coli XL-1-blue. Positive clones were identified by plasmid preparation and restriction analysis. The correct orientation of the metK fragment in pHS1 was determined by restriction digestion and sequencing. The resulting construct was designated pHS1 metK (Figure 1). The sequence of pHS1 metK is given in SEQ ID No.9. SEQ ID No.10 shows the amino acid sequence which is deduced from the metK-encoding region.

2. Construction of plasmids pHBbio14 and pHS1 bioS1

The construction of plasmids pHBbio14 and pHS1 bioS1 has already been described (DE 197.31274.8, Priority 22.7.97, Examples 1, 2 and 5).

3. Construction of pHS1 metK bioS1

The plasmids pHS1 bioS1 [SEQ ID No.11, (DE 197.31274.8, Priority 22.7.97), SEQ ID No.12 shows the amino acid sequence which is deduced from the bioS1-encoding region] and pHS1 metK (SEQ ID No.9) were purified using a plasmid preparation method (Boehringer). The fragment carrying the metK gene was isolated from pHS1 metK by digesting with Acc65I. pHS1 bioS1 was digested with Acc65I and dephosphorylated with shrimp alkaline phosphatase (SAP) (Boehringer Mannheim). After the SAP had been denatured in accordance with the manufacturer's instructions, the vector and the metK fragment were ligated, in a molar ratio of 1:3, using the Rapid DNA Ligation Kit in accordance with the manufacturer's instructions. The ligation mixture was transformed into strain E. coli XL-1-blue. Positive clones were identified by plasmid preparation and restriction analysis. The correct orientation of the metK fragment in pHS1 bioS1 was determined by means of restriction digestion and sequencing. The resulting construct was designated pHS1 metK bioS1 (Figure 2). The sequence of pHS1 metK bioS1 is given in SEQ ID No.13. SEQ ID No.14 shows the amino acid sequence which was deduced from the metK-encoding region; SEQ ID

No.15 shows the amino acid sequence which was deduced from the bioS1-encoding region.

4. Increasing biotin productivity by overexpressing metK, bioS1 and metK in combination with bioS1.

Spontaneously rifampicin-resistant colonies were isolated from strain BM4086 (Ketner and Campbell J. Molec. Biology 1975 96:13) by plating on rifampicin plates. A P1 lysate was generated from one of these resistant strains. The strain W3110 was transduced with this P1 lysate and clones were subsequently selected using rifampicin. The resulting strain was transformed with plasmid pHBbio14 using the $CaCl_2$ method (Maniatis et al. Molecular Cloning Cols Spring Harbour Laboratory Press 1989) and grown on LB containing 100 µg of ampicillin/ml. The isolated, transformed strain (LU5560) was in each case transformed with plasmid pHS1, pHS1 metK, pHS1 bioS1 or pHS1 metK bioS1 using the $CaCl_2$ method and then selected on LB agar containing 100 µg of ampicillin/ml and 25 µg of kanamycin/ml.

One colony from each of the transformants was in each case inoculated into a DYT culture containing the appropriate antibiotics and incubated for 12 h. The overnight culture (= ONC) was used to inoculate a 10 ml culture in TB medium (Sambrook, J. Fritsch, E F. Maniatis, T. 2nd ed. Cold Spring Harbor Laboratory Press., 1989 ISBN 0-87969-373-8), which contained 30 g of glycerol/l and the appropriate antibiotics. In the cases where plasmids pHS1, pHS1 metK, pHS1bioS1 and pHS1 metK bioS1 were present, 1mM IPTG and 0.5% arabinose were added simultaneously in order to induce expression of the metK and bioS1 genes or, respectively, the combination of the two genes. After 24 h, the cells were separated off from the culture supernatant by centrifugation and the biotin concentration in the supernatant was determined by means of a competitive ELISA employing streptavidin. The results of this determination are shown in Table I.

Table I: Determination of the biotin concentration

| Strain | Plasmid I | Plasmid II | Biotin mg/l |
|---|---|---|---|
| LU5580 | pHBbio14 | Control, without plasmid | 11 |
| LU5580 | pHBbio14 | pHS1 | 25 |
| LU5580 | pHBbio14 | pHS1 bioS1 | 45 |
| LU5580 | pHBbio14 | pHS1 metK | 37 |
| LU5580 | pHBbio14 | pHS1 metK bioS1 | 52 |

We claim:

1. A process for producing biotin wherein an S-adenosylmethionine synthase gene, having the sequence SEQ ID No. 1, and at least one further biotin biosynthesis gene bioS1, bioS2 or bioS3, having the sequences SEQ ID No. 3, SEQ ID No. 5 or SEQ ID No. 7, and also their functional variants, analogues or derivatives, are expressed in a prokaryotic or eukaryotic host organism which is able to synthesize biotin, this organism is cultured and the synthesized biotin is used directly after separating off the biomass or after purifying the biotin.

2. A process as claimed in claim 1, wherein the variants of the genes having the sequences SEQ ID No.1, SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7 are genes which, on the amino acid level deduced from the sequences as claimed in claim 1, exhibit a homology of from 30 to 100% and enable an increased synthesis of biotin to be achieved.

3. A process as claimed in claim 1 or 2, wherein an organism selected from the group of the genera Escherichia, Citrobacter, Serratia, Klebsiella, Salmonella, Pseudomonas, Comamonas, Acinetobacter, Azotobacter, Chromobacterium, Bacillus, Clostridium, Arthrobacter, Corynebacterium, Brevibacterium, Lactococcus, Lactobacillus, Streptomyces, Rhizobium, Agrobacterium, Staphylococcus, Rhodotorula, Sporobolomyces, Yarrowia, Schizosaccharomyces or Saccharomyces is used as the host organism.

4. A process as claimed in any of claims 1 to 3, wherein a regulation-defective biotin mutant is used as the host organism.

5. A process as claimed in any of claims 1 to 4, wherein at least one copy of the genes having the sequences SEQ ID No.1, SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7 as claimed in claim 1 is expressed in a prokaryotic or eukaryotic host organism either alone or together with one or more copies of at least one further biotin gene selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR.

6. A process as claimed in any of claims 1 to 5, wherein at least one copy of the genes having the sequences SEQ ID No.1, SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7 as claimed in claim 1 is expressed in a prokaryotic or eukaryotic host organism either alone or, on a shared vector or on separate vectors, together with one or more copies at least one further biotin gene selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR.

7. A gene construct which comprises an S-adenosylmethionine synthase gene, having the sequence SEQ ID No. 1, and at least one further biotin biosynthesis gene bioS1, bioS2 or bioS3, having the sequences SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7, and also their functional variants, analogues or derivatives, and which is functionally linked to one or more regulatory signals for the purpose of increasing gene expression and/or protein expression and/or whose natural regulation has been switched off.

8. A gene construct as claimed in claim 7, which has been inserted into a vector which is suitable for expressing the gene in a prokaryotic or eukaryotic host organism.

9. A gene construct as claimed in claim 7 or 8, wherein the genes having the sequences SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7, and also their functional variants, analogues or derivatives, are present in several copies in the gene construct.

10. A gene construct as claimed in any of claims 7 to 9, wherein the S-adenosylmethionine synthase gene, SEQ ID No. 1, and at least one further biotin biosynthesis gene bioS1, bioS2 or bioS3, having the sequences SEQ ID No. 3, SEQ ID No. 5 and SEQ ID No. 7, and also their functional variants, analogues or derivatives, as claimed in claim 7, are present in the gene construct or vector together with one or more copies of at least one further gene selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR.

11. An organism which comprises a gene construct as claimed in any of claims 7 to 10.

12. The use of the sequences as claimed in claim 1 for producing biotin.

13. The use of the bioS3 gene, having the sequence SEQ ID No. 7, or of its functional variants, analogues or derivatives, either alone or in combination with at least one further gene selected from the group S-adenosylmethionine synthase gene, bioS1, bioS2, bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR, for producing biotin.

14. The use of a gene construct as claimed in any of claims 7 to 10 for producing biotin.

Process for preparing biotin

Abstract of the disclosure

The invention relates to a gene construct which contains an S-adenosylmethionine synthase gene, having the sequence SEQ ID No. 1, and a biotin biosynthesis gene bioS1, bioS2 and/or bioS3, having the sequences SEQ ID No.3, SEQ ID No.5 and SEQ ID No.7, respectively, and, where appropriate, at least one further biotin synthesis gene sequence selected from the group bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY or bioR. The invention furthermore relates to organisms which contain this gene construct and to the use of the gene construct for preparing biotin, and also to a process for preparing biotin.

FIG.1

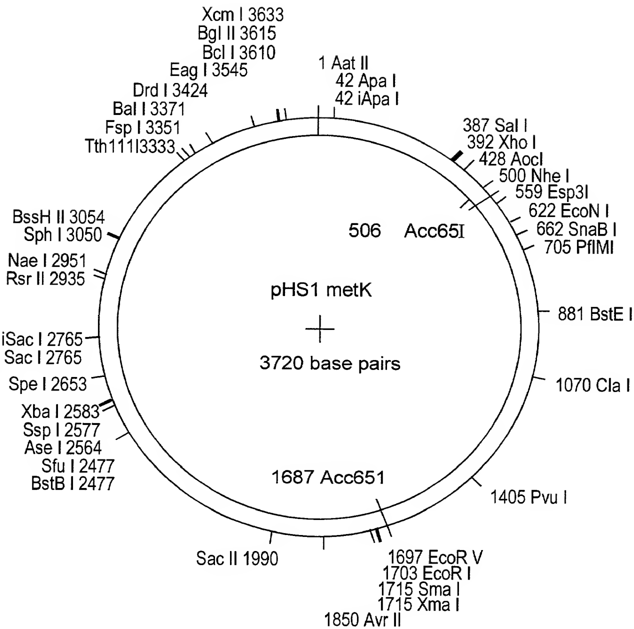
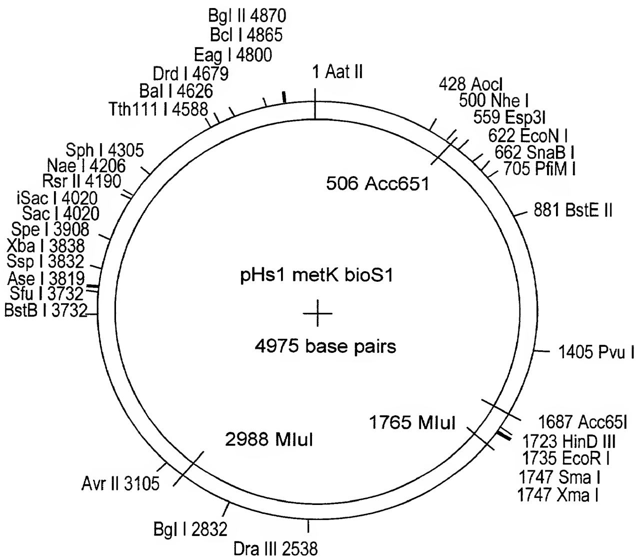

Xcm I 3633
Bgl II 3615
Bcl I 3610
Eag I 3545
Drd I 3424
Bal I 3371
Fsp I 3351
Tth111I3333
1 Aat II
42 Apa I
42 iApa I
387 Sal I
392 Xho I
428 AocI
500 Nhe I
559 Esp3I
622 EcoN I
662 SnaB I
705 PflMI
506 Acc65I
BssH II 3054
Sph I 3050
Nae I 2951
Rsr II 2935
pHS1 metK
3720 base pairs
881 BstE II
iSac I 2765
Sac I 2765
Spe I 2653
1070 Cla I
Xba I 2583
Ssp I 2577
Ase I 2564
Sfu I 2477
BstB I 2477
1687 Acc651
1405 Pvu I
Sac II 1990
1697 EcoR V
1703 EcoR I
1715 Sma I
1715 Xma I
1850 Avr II

FIG.2


Bgl II 4870
Bcl I 4865
Eag I 4800
Drd I 4679
Bal I 4626
Tth111 I 4588
1 Aat II
428 AocI
500 Nhe I
559 Esp3I
622 EcoN I
662 SnaB I
705 PflM I
506 Acc651
881 BstE II
Sph I 4305
Nae I 4206
Rsr II 4190
iSac I 4020
Sac I 4020
Spe I 3908
Xba I 3838
Ssp I 3832
Ase I 3819
Sfu I 3732
BstB I 3732
pHs1 metK bioS1
4975 base pairs
1405 Pvu I
1765 MluI
2988 MluI
1687 Acc65I
1723 HinD III
1735 EcoR I
1747 Sma I
1747 Xma I
Avr II 3105
Bgl I 2832
Dra III 2538

09/622419
528 Rec'd PCT/PTO 16 AUG 2000

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: BASF Aktiengesellschaft
(B) STREET: Karl Bosch Strasse
(C) CITY: Ludwigshafen
(D) FEDERAL STATE: Rheinland-Pfalz
(E) COUNTRY: Germany
(F) POSTAL CODE: 67056

(ii) TITLE OF APPLICATION: Process for preparing biotin

(iii) NUMBER OF SEQUENCES: 15

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID No: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1155 Base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:
(B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:
(B) CLONE: metK

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1155

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 1:

```
ATG GCA AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT      48
Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
 1               5                  10                  15

GAC AAA ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC      96
Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
            20                  25                  30

GAA CAG GAT CCG AAA GCA CGC GTT GCT TGC GAA ACC TAC GTA AAA ACC     144
Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
        35                  40                  45

GGC ATG GTT TTA GTT GGC GGC GAA ATC ACC ACC AGC GCC TGG GTA GAC     192
Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
    50                  55                  60

ATC GAA GAG ATC ACC CGT AAC ACC GTT CGC GAA ATT GGC TAT GTG CAT     240
Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
65                  70                  75                  80

TCC GAC ATG GGC TTT GAC GCT AAC TCC TGT GCG GTT CTG AGC GCT ATC     288
Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
                85                  90                  95

GGC AAA CAG TCT CCT GAC ATC AAC CAG GGC GTT GAC CGT GCC GAT CCG     336
Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
            100                 105                 110

CTG GAA CAG GGC GCG GGT GAC CAG GGT CTG ATG TTT GGC TAC GCA ACT     384
Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
        115                 120                 125

AAT GAA ACC GAC GTG CTG ATG CCA GCA CCT ATC ACC TAT GCA CAC CGT     432
Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
    130                 135                 140

CTG GTA CAG CGT CAG GCT GAA GTG CGT AAA AAC GGC ACT CTG CCG TGG     480
Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145                 150                 155                 160

CTG CGC CCG GAC GCG AAA AGC CAG GTG ACT TTT CAG TAT GAC GAC GGC     528
Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
                165                 170                 175

AAA ATC GTT GGT ATC GAT GCT GTC GTG CTT TCC ACT CAG CAC TCT GAA     576
Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
            180                 185                 190
```

GAG ATC GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC 624
Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

AAG CCA ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC 672
Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

ATC AAC CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC 720
Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

GGT CTG ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG 768
Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

CGT CAC GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC 816
Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

CGT TCC GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT 864
Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

GCT GGC CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC 912
Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

GTG GCT GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA 960
Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

GTG CCT TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG 1008
Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

CGC CCA TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC 1056
Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

AAA GAA ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG 1104
Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

GAA AAA ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG 1152
Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

TAA 1155

385

(2) INFORMATION FOR SEQ ID No: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 2:

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
1 5 10 15

Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
20 25 30

Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
35 40 45

Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
50 55 60

Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
65 70 75 80

Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
85 90 95

Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
100 105 110

Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
115 120 125

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
130 135 140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
165 170 175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
180 185 190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

(2) INFORMATION FOR SEQ ID No: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1206 Base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:
(B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:
(B) CLONE: bioS1

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1206

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 3:

```
ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA        48
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
  1               5                  10                  15

CAG GAT GCG GGC GTC TAT CTC GAC AGC GCC GCG ACC GCG CTT AAA CCT        96
Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
             20                  25                  30

GAA GCC GTG GTT GAA GCC ACC CAA CAG TTT TAC AGT CTG AGC GCC GGA       144
Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
         35                  40                  45

AAC GTC CAT CGC AGC CAG TTT GCC GAA GCC CAA CGC CTG ACC GCG CGT       192
Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
     50                  55                  60

TAT GAA GCT GCA CGA GAG AAA GTG GCG CAA TTA CTG AAT GCA CCG GAT       240
Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
 65                  70                  75                  80

GAT AAA ACT ATC GTC TGG ACG CGC GGC ACC ACT GAA TCC ATC AAC ATG       288
Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
                 85                  90                  95

GTG GCA CAA TGC TAT GCG CGT CCG CGT CTG CAA CCG GGC GAT GAG ATT       336
Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
            100                 105                 110

ATT GTC AGC GTG GCA GAA CAC CAC GCC AAC CTC GTC CCC TGG CTG ATG       384
Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
        115                 120                 125

GTC GCC CAA CAA ACT GGA GCC AAA GTG GTG AAA TTG CCG CTT AAT GCG       432
Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
    130                 135                 140
```

CAG CGA CTG CCG GAT GTC GAT TTG TTG CCA GAA CTG ATT ACT CCC CGT 480
Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

AGT CGG ATT CTG GCG TTG GGT CAG ATG TCG AAC GTT ACT GGC GGT TGC 528
Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

CCG GAT CTG GCG CGA GCG ATT ACC TTT GCT CAT TCA GCC GGG ATG GTG 576
Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

GTG ATG GTT GAT GGT GCT CAG GGG GCA GTG CAT TTC CCC GCG GAT GTT 624
Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

CAG CAA CTG GAT ATT GAT TTC TAT GCT TTT TCA GGT CAC AAA CTG TAT 672
Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

GGC CCG ACA GGT ATC GGC GTG CTG TAT GGT AAA TCA GAA CTG CTG GAG 720
Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

GCG ATG TCG CCC TGG CTG GGC GGC GGC AAA ATG GTT CAC GAA GTG AGT 768
Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

TTT GAC GGC TTC ACG ACT CAA TCT GCG CCG TGG AAA CTG GAA GCT GGA 816
Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

ACG CCA AAT GTC GCT GGT GTC ATA GGA TTA AGC GCG GCG CTG GAA TGG 864
Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

CTG GCA GAT TAC GAT ATC AAC CAG GCC GAA AGC TGG AGC CGT AGC TTA 912
Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

GCA ACG CTG GCG GAA GAT GCG CTG GCG AAA CGT CCC GGC TTT CGT TCA 960
Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

TTC CGC TGC CAG GAT TCC AGC CTG CTG GCC TTT GAT TTT GCT GGC GTT 1008
Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

CAT CAT AGC GAT ATG GTG ACG CTG CTG GCG GAG TAC GGT ATT GCC CTG 1056
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

CGG GCC GGG CAG CAT TGC GCT CAG CCG CTA CTG GCA GAA TTA GGC GTA 1104
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

ACC GGC ACA CTG CGC GCC TCT TTT GCG CCA TAT AAT ACA AAG AGT GAT 1152
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

GTG GAT GCG CTG GTG AAT GCC GTT GAC CGC GCG CTG GAA TTA TTG GTG 1200
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

GAT TAA 1206
Asp

(2) INFORMATION FOR SEQ ID No: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 401 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 4:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
    370                 375                 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385                 390                 395                 400

Asp

(2) INFORMATION FOR SEQ ID No: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1215 Base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORGINAL SOURCE:
(B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:
(B) CLONE: bioS2

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1215

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 5:

ATG AAA TTA CCG ATT TAT CTC GAC TAC TCC GCA ACC ACG CCG GTG GAC 48
Met Lys Leu Pro Ile Tyr Leu Asp Tyr Ser Ala Thr Thr Pro Val Asp
  1               5                  10                  15

CCG CGT GTT GCC GAG AAA ATG ATG CAG TTT ATG ACG ATG GAC GGA ACC 96
Pro Arg Val Ala Glu Lys Met Met Gln Phe Met Thr Met Asp Gly Thr
            20                  25                  30

TTT GGT AAC CCG GCC TCC CGT TCT CAC CGT TTC GGC TGG CAG GCT GAA 144
Phe Gly Asn Pro Ala Ser Arg Ser His Arg Phe Gly Trp Gln Ala Glu
        35                  40                  45

GAA GCG GTA GAT ATC GCC CGT AAT CAG ATT GCC GAT CTG GTC GGC GCT 192
Glu Ala Val Asp Ile Ala Arg Asn Gln Ile Ala Asp Leu Val Gly Ala
    50                  55                  60

GAT CCG CGT GAA ATC GTC TTT ACC TCT GGT GCA ACC GAA TCT GAC AAC 240
Asp Pro Arg Glu Ile Val Phe Thr Ser Gly Ala Thr Glu Ser Asp Asn
65 70 75 80

CTG GCG ATC AAA GGT GCA GCC AAC TTT TAT CAG AAA AAA GGC AAG CAC 288
Leu Ala Ile Lys Gly Ala Ala Asn Phe Tyr Gln Lys Lys Gly Lys His
85 90 95

ATC ATC ACC AGC AAA ACC GAA CAC AAA GCG GTA CTG GAT ACC TGC CGT 336
Ile Ile Thr Ser Lys Thr Glu His Lys Ala Val Leu Asp Thr Cys Arg
100 105 110

CAG CTG GAG CGC GAA GGT TTT GAA GTC ACC TAC CTG GCA CCG CAG CGT 384
Gln Leu Glu Arg Glu Gly Phe Glu Val Thr Tyr Leu Ala Pro Gln Arg
115 120 125

AAC GGC ATT ATC GAC CTG AAA GAA CTT GAA GCA GCG ATG CGT GAC GAC 432
Asn Gly Ile Ile Asp Leu Lys Glu Leu Glu Ala Ala Met Arg Asp Asp
130 135 140

ACC ATC CTC GTG TCC ATC ATG CAC GTA AAT AAC GAA ATC GGC GTG GTG 480
Thr Ile Leu Val Ser Ile Met His Val Asn Asn Glu Ile Gly Val Val
145 150 155 160

CAG GAT ATC GCG GCT ATC GGC GAA ATG TGC CGT GCT CGT GGC ATT ATC 528
Gln Asp Ile Ala Ala Ile Gly Glu Met Cys Arg Ala Arg Gly Ile Ile
165 170 175

TAT CAC GTT GAT GCA ACC CAG AGC GTG GGT AAA CTG CCT ATC GAC CTG 576
Tyr His Val Asp Ala Thr Gln Ser Val Gly Lys Leu Pro Ile Asp Leu
180 185 190

AGC CAG TTG AAA GTT GAC CTG ATG TCT TTC TCC GGT CAC AAA ATC TAT 624
Ser Gln Leu Lys Val Asp Leu Met Ser Phe Ser Gly His Lys Ile Tyr
195 200 205

GGC CCG AAA GGT ATC GGT GCG CTG TAT GTA CGT CGT AAA CCG CGC GTA 672
Gly Pro Lys Gly Ile Gly Ala Leu Tyr Val Arg Arg Lys Pro Arg Val
210 215 220

CGC ATC GAA GCG CAA ATG CAC GGC GGC GGT CAC GAG CGC GGT ATG CGT 720
Arg Ile Glu Ala Gln Met His Gly Gly Gly His Glu Arg Gly Met Arg
225 230 235 240

TCC GGC ACT CTG CCT GTT CAC CAG ATC GTC GGA ATG GGC GAG GCC TAT 768
Ser Gly Thr Leu Pro Val His Gln Ile Val Gly Met Gly Glu Ala Tyr
245 250 255

CGC ATC GCA AAA GAA GAG ATG GCG ACC GAG ATG GAA CGT CTG CGC GGC 816
Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly
260 265 270

CTG CGT AAC CGT CTG TGG AAC GGC ATC AAA GAT ATC GAA GAA GTT TAC 864
Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr
275 280 285

CTG AAC GGT GAC CTG GAA CAC GGT GCG CCG AAC ATT CTC AAC GTC AGC 912
Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser
290 295 300

TTC AAC TAC GTT GAA GGT GAG TCG CTG ATT ATG GCG CTG AAA GAC CTC 960
Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu
305 310 315 320

GCA GTT TCT TCA GGT TCC GCC TGT ACG TCA GCA AGC CTC GAA CCG TCC 1008
Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser
325 330 335

TAC GTG CTG CGC GCG CTG GGG CTG AAC GAC GAG CTG GCA CAT AGC TCT 1056
Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser
340 345 350

ATC CGT TTC TCT TTA GGT CGT TTT ACT ACT GAA GAA GAG ATC GAC TAC 1104
Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr
355 360 365

ACC ATC GAG TTA GTT CGT AAA TCC ATC GGT CGT CTG CGT GAC CTT TCT 1152
Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser
370 375 380

CCG CTG TGG GAA ATG TAC AAG CAG GGC GTG GAT CTG AAC AGC ATC GAA 1200
Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu
385 390 395 400

TGG GCT CAT CAT TAA 1215
Trp Ala His His
405

(2) INFORMATION FOR SEQ ID No: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 404 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 6:

```
Met Lys Leu Pro Ile Tyr Leu Asp Tyr Ser Ala Thr Thr Pro Val Asp
 1               5                  10                  15

Pro Arg Val Ala Glu Lys Met Met Gln Phe Met Thr Met Asp Gly Thr
            20                  25                  30

Phe Gly Asn Pro Ala Ser Arg Ser His Arg Phe Gly Trp Gln Ala Glu
        35                  40                  45

Glu Ala Val Asp Ile Ala Arg Asn Gln Ile Ala Asp Leu Val Gly Ala
    50                  55                  60

Asp Pro Arg Glu Ile Val Phe Thr Ser Gly Ala Thr Glu Ser Asp Asn
65                  70                  75                  80

Leu Ala Ile Lys Gly Ala Ala Asn Phe Tyr Gln Lys Lys Gly Lys His
                85                  90                  95

Ile Ile Thr Ser Lys Thr Glu His Lys Ala Val Leu Asp Thr Cys Arg
            100                 105                 110

Gln Leu Glu Arg Glu Gly Phe Glu Val Thr Tyr Leu Ala Pro Gln Arg
        115                 120                 125

Asn Gly Ile Ile Asp Leu Lys Glu Leu Glu Ala Ala Met Arg Asp Asp
    130                 135                 140

Thr Ile Leu Val Ser Ile Met His Val Asn Asn Glu Ile Gly Val Val
145                 150                 155                 160

Gln Asp Ile Ala Ala Ile Gly Glu Met Cys Arg Ala Arg Gly Ile Ile
                165                 170                 175

Tyr His Val Asp Ala Thr Gln Ser Val Gly Lys Leu Pro Ile Asp Leu
            180                 185                 190

Ser Gln Leu Lys Val Asp Leu Met Ser Phe Ser Gly His Lys Ile Tyr
        195                 200                 205

Gly Pro Lys Gly Ile Gly Ala Leu Tyr Val Arg Arg Lys Pro Arg Val
    210                 215                 220

Arg Ile Glu Ala Gln Met His Gly Gly Gly His Glu Arg Gly Met Arg
225                 230                 235                 240

Ser Gly Thr Leu Pro Val His Gln Ile Val Gly Met Gly Glu Ala Tyr
                245                 250                 255
```

Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly
260 265 270

Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr
275 280 285

Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser
290 295 300

Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu
305 310 315 320

Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser
325 330 335

Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser
340 345 350

Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr
355 360 365

Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser
370 375 380

Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu
385 390 395 400

Trp Ala His His

(2) INFORMATION FOR SEQ ID No: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1221 Base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNS (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:
(B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:
(B) CLONE: bioS3

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1221

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 7:

```
ATG ATT TTT TCC GTC GAC AAA GTG CGG GCC GAC TTT CCG GTG CTT TCG       48
Met Ile Phe Ser Val Asp Lys Val Arg Ala Asp Phe Pro Val Leu Ser
  1               5                  10                  15

CGT GAG GTA AAC GGT TTG CCG CTG GCT TAT CTC GAC AGC GCC GCC AGT       96
Arg Glu Val Asn Gly Leu Pro Leu Ala Tyr Leu Asp Ser Ala Ala Ser
             20                  25                  30

GCG CAG AAA CCG AGC CAG GTG ATT GAC GCC GAG GCC GAG TTT TAT CGT      144
Ala Gln Lys Pro Ser Gln Val Ile Asp Ala Glu Ala Glu Phe Tyr Arg
         35                  40                  45

CAT GGC TAC GCG GCG GTG CAT CGT GGT ATT CAT ACC TTA AGC GCC CAG      192
His Gly Tyr Ala Ala Val His Arg Gly Ile His Thr Leu Ser Ala Gln
     50                  55                  60

GCG ACC GAG AAA ATG GAG AAC GTG CGC AAG CGG GCA TCG CTG TTT ATT      240
Ala Thr Glu Lys Met Glu Asn Val Arg Lys Arg Ala Ser Leu Phe Ile
 65                  70                  75                  80

AAT GCC CGT TCG GCG GAA GAG CTG GTG TTC GTC CGC GGC ACG ACG GAA      288
Asn Ala Arg Ser Ala Glu Glu Leu Val Phe Val Arg Gly Thr Thr Glu
                 85                  90                  95

GGG ATC AAT CTG GTC GCC AAT AGC TGG GGC AAC AGC AAC GTG CGG GCG      336
Gly Ile Asn Leu Val Ala Asn Ser Trp Gly Asn Ser Asn Val Arg Ala
            100                 105                 110

GGC GAT AAC ATC ATC ATC AGT CAG ATG GAG CAC CAC GCT AAC ATT GTT      384
Gly Asp Asn Ile Ile Ile Ser Gln Met Glu His His Ala Asn Ile Val
        115                 120                 125

CCC TGG CAG ATG CTT TGC GCA CGC GTT GGC GCA GAG CTG CGT GTG ATC      432
Pro Trp Gln Met Leu Cys Ala Arg Val Gly Ala Glu Leu Arg Val Ile
    130                 135                 140

CCG CTC AAT CCC GAT GGT ACG TTG CAA CTG GAG ACG CTG CCT ACG CTG      480
Pro Leu Asn Pro Asp Gly Thr Leu Gln Leu Glu Thr Leu Pro Thr Leu
145                 150                 155                 160

TTT GAT GAG AAA ACT CGC CTG CTG GCA ATT ACT CAT GTC TCC AAC GTG      528
Phe Asp Glu Lys Thr Arg Leu Leu Ala Ile Thr His Val Ser Asn Val
                165                 170                 175
```

CTT GGC ACA GAA AAT CCA CTG GCG GAA ATG ATC ACG CTT GCG CAC CAG 576
Leu Gly Thr Glu Asn Pro Leu Ala Glu Met Ile Thr Leu Ala His Gln
180 185 190

CAT GGC GCA AAA GTG CTG GTG GAT GGC GCT CAG GCG GTG ATG CAT CAT 624
His Gly Ala Lys Val Leu Val Asp Gly Ala Gln Ala Val Met His His
195 200 205

CCG GTG GAT GTT CAG GCG CTG GAT TGC GAC TTT TAC GTG TTC TCC GGG 672
Pro Val Asp Val Gln Ala Leu Asp Cys Asp Phe Tyr Val Phe Ser Gly
210 215 220

CAT AAA CTG TAT GGC CCC ACC GGA ATT GGC ATT CTT TAT GTG AAA GAA 720
His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Ile Leu Tyr Val Lys Glu
225 230 235 240

GCC TTG TTG CAG GAG ATG CCG CCG TGG GAA GGG GGC GGT TCT ATG ATC 768
Ala Leu Leu Gln Glu Met Pro Pro Trp Glu Gly Gly Gly Ser Met Ile
245 250 255

GCC ACC GTC AGC CTG AGT GAA GGC ACT ACC TGG ACC AAA GCA CCA TGG 816
Ala Thr Val Ser Leu Ser Glu Gly Thr Thr Trp Thr Lys Ala Pro Trp
260 265 270

CGG TTT GAA GCC GGT ACA CCC AAT ACC GGG GGC ATC ATT GGT CTT GGC 864
Arg Phe Glu Ala Gly Thr Pro Asn Thr Gly Gly Ile Ile Gly Leu Gly
275 280 285

GCG GCG CTG GAG TAT GTT TCG GCG CTG GGG CTT AAT AAC ATA GCC GAG 912
Ala Ala Leu Glu Tyr Val Ser Ala Leu Gly Leu Asn Asn Ile Ala Glu
290 295 300

TAT GAA CAG AAT CTG ATG CAT TAT GCG CTA TCA CAG CTG GAA TCT GTA 960
Tyr Glu Gln Asn Leu Met His Tyr Ala Leu Ser Gln Leu Glu Ser Val
305 310 315 320

CCG GAT CTC ACT CTC TAT GGC CCA CAA AAC AGG CTT GGC GTT ATT GCT 1008
Pro Asp Leu Thr Leu Tyr Gly Pro Gln Asn Arg Leu Gly Val Ile Ala
325 330 335

TTT AAT CTC GGT AAA CAC CAC GCC TAT GAT GTT GGC AGT TTT CTC GAT 1056
Phe Asn Leu Gly Lys His His Ala Tyr Asp Val Gly Ser Phe Leu Asp
340 345 350

AAT TAC GGC ATT GCT GTG CGT ACC GGA CAT CAC TGC GCA ATG CCA TTG 1104
Asn Tyr Gly Ile Ala Val Arg Thr Gly His His Cys Ala Met Pro Leu
355 360 365

ATG GCC TAT TAC AAC GTC CCT GCG ATG TGT CGG GCG TCG CTG GCC ATG 1152
Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met
370 375 380

TAT AAC ACC CAT GAA GAA GTG GAT CGT CTG GTG ACC GGC CTG CAA CGT 1200
Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg
385 390 395 400

ATT CAC CGT TTG CTG GGA TAA 1221
Ile His Arg Leu Leu Gly
405

(2) INFORMATION FOR SEQ ID No: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 406 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 8:

Met Ile Phe Ser Val Asp Lys Val Arg Ala Asp Phe Pro Val Leu Ser
1 5 10 15

Arg Glu Val Asn Gly Leu Pro Leu Ala Tyr Leu Asp Ser Ala Ala Ser
20 25 30

Ala Gln Lys Pro Ser Gln Val Ile Asp Ala Glu Ala Glu Phe Tyr Arg
35 40 45

His Gly Tyr Ala Ala Val His Arg Gly Ile His Thr Leu Ser Ala Gln
50 55 60

Ala Thr Glu Lys Met Glu Asn Val Arg Lys Arg Ala Ser Leu Phe Ile
65 70 75 80

Asn Ala Arg Ser Ala Glu Glu Leu Val Phe Val Arg Gly Thr Thr Glu
85 90 95

Gly Ile Asn Leu Val Ala Asn Ser Trp Gly Asn Ser Asn Val Arg Ala
100 105 110

Gly Asp Asn Ile Ile Ile Ser Gln Met Glu His His Ala Asn Ile Val
115 120 125

Pro Trp Gln Met Leu Cys Ala Arg Val Gly Ala Glu Leu Arg Val Ile
130 135 140

```
Pro Leu Asn Pro Asp Gly Thr Leu Gln Leu Glu Thr Leu Pro Thr Leu
145                 150                 155                 160

Phe Asp Glu Lys Thr Arg Leu Leu Ala Ile Thr His Val Ser Asn Val
                165                 170                 175

Leu Gly Thr Glu Asn Pro Leu Ala Glu Met Ile Thr Leu Ala His Gln
            180                 185                 190

His Gly Ala Lys Val Leu Val Asp Gly Ala Gln Ala Val Met His His
        195                 200                 205

Pro Val Asp Val Gln Ala Leu Asp Cys Asp Phe Tyr Val Phe Ser Gly
    210                 215                 220

His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Ile Leu Tyr Val Lys Glu
225                 230                 235                 240

Ala Leu Leu Gln Glu Met Pro Pro Trp Glu Gly Gly Gly Ser Met Ile
                245                 250                 255

Ala Thr Val Ser Leu Ser Glu Gly Thr Thr Trp Thr Lys Ala Pro Trp
            260                 265                 270

Arg Phe Glu Ala Gly Thr Pro Asn Thr Gly Gly Ile Ile Gly Leu Gly
        275                 280                 285

Ala Ala Leu Glu Tyr Val Ser Ala Leu Gly Leu Asn Asn Ile Ala Glu
    290                 295                 300

Tyr Glu Gln Asn Leu Met His Tyr Ala Leu Ser Gln Leu Glu Ser Val
305                 310                 315                 320

Pro Asp Leu Thr Leu Tyr Gly Pro Gln Asn Arg Leu Gly Val Ile Ala
                325                 330                 335

Phe Asn Leu Gly Lys His His Ala Tyr Asp Val Gly Ser Phe Leu Asp
            340                 345                 350

Asn Tyr Gly Ile Ala Val Arg Thr Gly His His Cys Ala Met Pro Leu
        355                 360                 365

Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met
    370                 375                 380

Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg
385                 390                 395                 400

Ile His Arg Leu Leu Gly
                405
```

(2) INFORMATION FOR SEQ ID No: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3720 Base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: pHS1 metK

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 530..1684

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 9:

```
GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTCACAC AGGGCCCTCG GACACCGAGG       60

AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC      120

GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA      180

AGGTAGTTAT CCTTAAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA      240

AGCTCATTTC TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG      300

AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA      360

ATGTCCTGAG CAATCACCTA TGAACTGTCG ACTCGAGATA GCATTTTTAT CCATAAGATT      420

AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC      480

GGATAACAAT TTCACACACG CTAGCGGTAC CAAAGAGGAG AAATTAACT ATG GCA          535
                                                      Met Ala
                                                       1

AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA        583
Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys
         5                  10                  15

ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG        631
Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln
     20                  25                  30
```

GAT CCG AAA GCA CGC GTT GCT TGC GAA ACC TAC GTA AAA ACC GGC ATG 679
Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr Gly Met
35 40 45 50

GTT TTA GTT GGC GGC GAA ATC ACC ACC AGC GCC TGG GTA GAC ATC GAA 727
Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp Ile Glu
55 60 65

GAG ATC ACC CGT AAC ACC GTT CGC GAA ATT GGC TAT GTG CAT TCC GAC 775
Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His Ser Asp
70 75 80

ATG GGC TTT GAC GCT AAC TCC TGT GCG GTT CTG AGC GCT ATC GGC AAA 823
Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile Gly Lys
85 90 95

CAG TCT CCT GAC ATC AAC CAG GGC GTT GAC CGT GCC GAT CCG CTG GAA 871
Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro Leu Glu
100 105 110

CAG GGC GCG GGT GAC CAG GGT CTG ATG TTT GGC TAC GCA ACT AAT GAA 919
Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr Asn Glu
115 120 125 130

ACC GAC GTG CTG ATG CCA GCA CCT ATC ACC TAT GCA CAC CGT CTG GTA 967
Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg Leu Val
135 140 145

CAG CGT CAG GCT GAA GTG CGT AAA AAC GGC ACT CTG CCG TGG CTG CGC 1015
Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp Leu Arg
150 155 160

CCG GAC GCG AAA AGC CAG GTG ACT TTT CAG TAT GAC GAC GGC AAA ATC 1063
Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly Lys Ile
165 170 175

GTT GGT ATC GAT GCT GTC GTG CTT TCC ACT CAG CAC TCT GAA GAG ATC 1111
Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu Glu Ile
180 185 190

GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC AAG CCA 1159
Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile Lys Pro
195 200 205 210

ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC ATC AAC 1207
Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe Ile Asn
215 220 225

CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG 1255
Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu
230 235 240

ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC 1303
Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His
245 250 255

GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC 1351
Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser
260 265 270

GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC 1399
Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly
275 280 285 290

CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT 1447
Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala
295 300 305

GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT 1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro
310 315 320

TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA 1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro
325 330 335

TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA 1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu
340 345 350

ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA 1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys
355 360 365 370

ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC 1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
375 380 385

CGCTTGATAT CGAATTCCTG CAGCCCGGGG GATCCCATGG TACGCGTGCT AGAGGCATCA 1751

AATAAAACGA AAGGCTCAGT CGAAAGACTG GGCCTTTCGT TTTATCTGTT GTTTGTCGGT 1811

GAACGCTCTC CTGAGTAGGA CAAATCCGCC GCCCTAGACC TAGGGGATAT ATTCCGCTTC 1871

CTCGCTCACT GACTCGCTAC GCTCGGTCGT TCGACTGCGG CGAGCGGAAA TGGCTTACGA 1931

ACGGGGCGGA GATTTCCTGG AAGATGCCAG GAAGATACTT AACAGGGAAG TGAGAGGGCC 1991

GCGGCAAAGC CGTTTTTCCA TAGGCTCCGC CCCCCTGACA AGCATCACGA AATCTGACGC 2051

TCAAATCAGT GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGC 2111

GGCTCCCTCG TGCGCTCTCC TGTTCCTGCC TTTCGGTTTA CCGGTGTCAT TCCGCTGTTA 2171

TGGCCGCGTT TGTCTCATTC CACGCCTGAC ACTCAGTTCC GGGTAGGCAG TTCGCTCCAA 2231

GCTGGACTGT ATGCACGAAC CCCCCGTTCA GTCCGACCGC TGCGCCTTAT CCGGTAACTA 2291

TCGTCTTGAG TCCAACCCGG AAAGACATGC AAAAGCACCA CTGGCAGCAG CCACTGGTAA 2351

TTGATTTAGA GGAGTTAGTC TTGAAGTCAT GCGCCGGTTA AGGCTAAACT GAAAGGACAA 2411

GTTTTGGTGA CTGCGCTCCT CCAAGCCAGT TACCTCGGTT CAAAGAGTTG GTAGCTCAGA 2471

GAACCTTCGA AAAACCGCCC TGCAAGGCGG TTTTTTCGTT TTCAGAGCAA GAGATTACGC 2531

GCAGACCAAA ACGATCTCAA GAAGATCATC TTATTAATCA GATAAAATAT TTCTAGATTT 2591

CAGTGCAATT TATCTCTTCA AATGTAGCAC CTGAAGTCAG CCCCATACGA TATAAGTTGT 2651

TACTAGTGCT TGGATTCTCA CCAATAAAAA ACGCCCGGCG GCAACCGAGC GTTCTGAACA 2711

AATCCAGATG GAGTTCTGAG GTCATTACTG GATCTATCAA CAGGAGTCCA AGCGAGCTCT 2771

CGAACCCCAG AGTCCCGCTC AGAAGAACTC GTCAAGAAGG CGATAGAAGG CGATGCGCTG 2831

CGAATCGGGA GCGGCGATAC CGTAAAGCAC GAGGAAGCGG TCAGCCCATT CGCCGCCAAG 2891

CTCTTCAGCA ATATCACGGG TAGCCAACGC TATGTCCTGA TAGCGGTCCG CCACACCCAG 2951

CCGGCCACAG TCGATGAATC CAGAAAAGCG GCCATTTTCC ACCATGATAT TCGGCAAGCA 3011

GGCATCGCCA TGGGTCACGA CGAGATCCTC GCCGTCGGGC ATGCGCGCCT TGAGCCTGGC 3071

GAACAGTTCG GCTGGCGCGA GCCCCTGATG CTCTTCGTCC AGATCATCCT GATCGACAAG 3131

ACCGGCTTCC ATCCGAGTAC GTGCTCGCTC GATGCGATGT TTCGCTTGGT GGTCGAATGG 3191

GCAGGTAGCC GGATCAAGCG TATGCAGCCG CCGCATTGCA TCAGCCATGA TGGATACTTT 3251

CTCGGCAGGA GCAAGGTGAG ATGACAGGAG ATCCTGCCCC GGCACTTCGC CCAATAGCAG 3311

CCAGTCCCTT CCCGCTTCAG TGACAACGTC GAGCACAGCT GCGCAAGGAA CGCCCGTCGT 3371

GGCCAGCCAC GATAGCCGCG CTGCCTCGTC CTGCAGTTCA TTCAGGGCAC CGGACAGGTC 3431

GGTCTTGACA AAAAGAACCG GGCGCCCCTG CGCTGACAGC CGGAACACGG CGGCATCAGA 3491

GCAGCCGATT GTCTGTTGTG CCCAGTCATA GCCGAATAGC CTCTCCACCC AAGCGGCCGG 3551

AGAACCTGCG TGCAATCCAT CTTGTTCAAT CATGCGAAAC GATCCTCATC CTGTCTCTTG 3611

ATCAGATCTT GATCCCCTGC GCCATCAGAT CCTTGGCGGC AAGAAAGCCA TCCAGTTTAC 3671

TTTGCAGGGC TTCCCAACCT TACCAGAGGG CGCCCCAGCT GGCAATTCC 3720

(2) INFORMATION FOR SEQ ID No: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 10:

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
1 5 10 15

Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
20 25 30

Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
35 40 45

Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
50 55 60

Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
65 70 75 80

Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
85 90 95

Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
100 105 110

Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
115 120 125

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
130 135 140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
165 170 175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
180 185 190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

(2) INFORMATION FOR SEQ ID No: 11:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3794 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: pHS1 bioS1

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 601..1806

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 11:

GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTCACAC AGGGCCCTCG GACACCGAGG 60

AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC 120

GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA 180

AGGTAGTTAT CCTTAAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA 240

AGCTCATTTC TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG 300

AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA 360

ATGTCCTGAG CAATCACCTA TGAACTGTCG ACTCGAGATA GCATTTTTAT CCATAAGATT 420

AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC 480

GGATAACAAT TTCACACACG CTAGCGGTAC CGGGCCCCCC CTCGAGGTCG ACGGTATCGA 540

TAAGCTTGAT ATCGAATTCC TGCAGCCCGG GGGATCCCAT GGTACGCGTC GAGGAGTACC 600

ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA 648
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15

CAG GAT GCG GGC GTC TAT CTC GAC AGC GCC GCG ACC GCG CTT AAA CCT 696
Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30

GAA GCC GTG GTT GAA GCC ACC CAA CAG TTT TAC AGT CTG AGC GCC GGA 744
Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45

AAC GTC CAT CGC AGC CAG TTT GCC GAA GCC CAA CGC CTG ACC GCG CGT 792
Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60

TAT GAA GCT GCA CGA GAG AAA GTG GCG CAA TTA CTG AAT GCA CCG GAT 840
Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80

GAT AAA ACT ATC GTC TGG ACG CGC GGC ACC ACT GAA TCC ATC AAC ATG 888
Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95

GTG GCA CAA TGC TAT GCG CGT CCG CGT CTG CAA CCG GGC GAT GAG ATT 936
Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110

ATT GTC AGC GTG GCA GAA CAC CAC GCC AAC CTC GTC CCC TGG CTG ATG 984
Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125

GTC GCC CAA CAA ACT GGA GCC AAA GTG GTG AAA TTG CCG CTT AAT GCG 1032
Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140

CAG CGA CTG CCG GAT GTC GAT TTG TTG CCA GAA CTG ATT ACT CCC CGT 1080
Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

AGT CGG ATT CTG GCG TTG GGT CAG ATG TCG AAC GTT ACT GGC GGT TGC 1128
Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

CCG GAT CTG GCG CGA GCG ATT ACC TTT GCT CAT TCA GCC GGG ATG GTG 1176
Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

GTG ATG GTT GAT GGT GCT CAG GGG GCA GTG CAT TTC CCC GCG GAT GTT 1224
Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

CAG CAA CTG GAT ATT GAT TTC TAT GCT TTT TCA GGT CAC AAA CTG TAT 1272
Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

GGC CCG ACA GGT ATC GGC GTG CTG TAT GGT AAA TCA GAA CTG CTG GAG 1320
Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

GCG ATG TCG CCC TGG CTG GGC GGC GGC AAA ATG GTT CAC GAA GTG AGT 1368
Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

TTT GAC GGC TTC ACG ACT CAA TCT GCG CCG TGG AAA CTG GAA GCT GGA 1416
Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

ACG CCA AAT GTC GCT GGT GTC ATA GGA TTA AGC GCG GCG CTG GAA TGG 1464
Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

CTG GCA GAT TAC GAT ATC AAC CAG GCC GAA AGC TGG AGC CGT AGC TTA 1512
Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

GCA ACG CTG GCG GAA GAT GCG CTG GCG AAA CGT CCC GGC TTT CGT TCA 1560
Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

TTC CGC TGC CAG GAT TCC AGC CTG CTG GCC TTT GAT TTT GCT GGC GTT 1608
Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

CAT CAT AGC GAT ATG GTG ACG CTG CTG GCG GAG TAC GGT ATT GCC CTG 1656
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

CGG GCC GGG CAG CAT TGC GCT CAG CCG CTA CTG GCA GAA TTA GGC GTA 1704
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

ACC GGC ACA CTG CGC GCC TCT TTT GCG CCA TAT AAT ACA AAG AGT GAT 1752
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

GTG GAT GCG CTG GTG AAT GCC GTT GAC CGC GCG CTG GAA TTA TTG GTG 1800
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

GAT TAAACGCGTG CTAGAGGCAT CAAATAAAAC GAAAGGCTCA GTCGAAAGAC 1853
Asp

TGGGCCTTTC GTTTTATCTG TTGTTTGTCG GTGAACGCTC TCCTGAGTAG GACAAATCCG 1913

CCGCCCTAGA CCTAGGGGAT ATATTCCGCT TCCTCGCTCA CTGACTCGCT ACGCTCGGTC 1973

GTTCGACTGC GGCGAGCGGA AATGGCTTAC GAACGGGGCG GAGATTTCCT GGAAGATGCC 2033

AGGAAGATAC TTAACAGGGA AGTGAGAGGG CCGCGGCAAA GCCGTTTTTC CATAGGCTCC 2093

GCCCCCCTGA CAAGCATCAC GAAATCTGAC GCTCAAATCA GTGGTGGCGA AACCCGACAG 2153

GACTATAAAG ATACCAGGCG TTTCCCCCTG GCGGCTCCCT CGTGCGCTCT CCTGTTCCTG 2213

CCTTTCGGTT TACCGGTGTC ATTCCGCTGT TATGGCCGCG TTTGTCTCAT TCCACGCCTG 2273

ACACTCAGTT CCGGGTAGGC AGTTCGCTCC AAGCTGGACT GTATGCACGA ACCCCCCGTT 2333

CAGTCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGAAAGACAT 2393

GCAAAAGCAC CACTGGCAGC AGCCACTGGT AATTGATTTA GAGGAGTTAG TCTTGAAGTC 2453

ATGCGCCGGT TAAGGCTAAA CTGAAAGGAC AAGTTTTGGT GACTGCGCTC CTCCAAGCCA 2513

GTTACCTCGG TTCAAAGAGT TGGTAGCTCA GAGAACCTTC GAAAAACCGC CCTGCAAGGC 2573

GGTTTTTTCG TTTTCAGAGC AAGAGATTAC GCGCAGACCA AAACGATCTC AAGAAGATCA 2633

TCTTATTAAT CAGATAAAAT ATTTCTAGAT TTCAGTGCAA TTTATCTCTT CAAATGTAGC 2693

ACCTGAAGTC AGCCCCATAC GATATAAGTT GTTACTAGTG CTTGGATTCT CACCAATAAA 2753

AAACGCCCGG CGGCAACCGA GCGTTCTGAA CAAATCCAGA TGGAGTTCTG AGGTCATTAC 2813

TGGATCTATC AACAGGAGTC CAAGCGAGCT CTCGAACCCC AGAGTCCCGC TCAGAAGAAC 2873

TCGTCAAGAA GGCGATAGAA GGCGATGCGC TGCGAATCGG GAGCGGCGAT ACCGTAAAGC 2933

ACGAGGAAGC GGTCAGCCCA TTCGCCGCCA AGCTCTTCAG CAATATCACG GGTAGCCAAC 2993

GCTATGTCCT GATAGCGGTC CGCCACACCC AGCCGGCCAC AGTCGATGAA TCCAGAAAAG 3053

CGGCCATTTT CCACCATGAT ATTCGGCAAG CAGGCATCGC CATGGGTCAC GACGAGATCC 3113

TCGCCGTCGG GCATGCGCGC CTTGAGCCTG GCGAACAGTT CGGCTGGCGC GAGCCCCTGA 3173

TGCTCTTCGT CCAGATCATC CTGATCGACA AGACCGGCTT CCATCCGAGT ACGTGCTCGC 3233

TCGATGCGAT GTTTCGCTTG GTGGTCGAAT GGGCAGGTAG CCGGATCAAG CGTATGCAGC 3293

CGCCGCATTG CATCAGCCAT GATGGATACT TTCTCGGCAG GAGCAAGGTG AGATGACAGG 3353

AGATCCTGCC CCGGCACTTC GCCCAATAGC AGCCAGTCCC TTCCCGCTTC AGTGACAACG 3413

TCGAGCACAG CTGCGCAAGG AACGCCCGTC GTGGCCAGCC ACGATAGCCG CGCTGCCTCG 3473

TCCTGCAGTT CATTCAGGGC ACCGGACAGG TCGGTCTTGA CAAAAAGAAC CGGGCGCCCC 3533

TGCGCTGACA GCCGGAACAC GGCGGCATCA GAGCAGCCGA TTGTCTGTTG TGCCCAGTCA 3593

TAGCCGAATA GCCTCTCCAC CCAAGCGGCC GGAGAACCTG CGTGCAATCC ATCTTGTTCA 3653

ATCATGCGAA ACGATCCTCA TCCTGTCTCT TGATCAGATC TTGATCCCCT GCGCCATCAG 3713

ATCCTTGGCG GCAAGAAAGC CATCCAGTTT ACTTTGCAGG GCTTCCCAAC CTTACCAGAG 3773

GGCGCCCCAG CTGGCAATTC C 3794

(2) INFORMATION FOR SEQ ID No: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 401 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 12:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

Asp

(2) INFORMATION FOR SEQ ID No: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4975 Base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Single
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: pHS1 metK bioS1

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 1782..2987

(ix) FEATURES:
(A) NAME/KEY: CDS
(B) LOCATION: 530..1684

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 13:

```
GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTCACAC AGGGCCCTCG GACACCGAGG      60

AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC     120

GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA     180

AGGTAGTTAT CCTTAAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA     240

AGCTCATTTC TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG     300

AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA     360

ATGTCCTGAG CAATCACCTA TGAACTGTCG ACTCGAGATA GCATTTTTAT CCATAAGATT     420

AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC     480

GGATAACAAT TTCACACACG CTAGCGGTAC CAAAGAGGAG AAATTAACT ATG GCA         535
                                                      Met Ala
                                                        1

AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA       583
Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys
          5                  10                  15

ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG       631
Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln
     20                  25                  30
```

GAT CCG AAA GCA CGC GTT GCT TGC GAA ACC TAC GTA AAA ACC GGC ATG 679
Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr Gly Met
35 40 45 50

GTT TTA GTT GGC GGC GAA ATC ACC ACC AGC GCC TGG GTA GAC ATC GAA 727
Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp Ile Glu
55 60 65

GAG ATC ACC CGT AAC ACC GTT CGC GAA ATT GGC TAT GTG CAT TCC GAC 775
Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His Ser Asp
70 75 80

ATG GGC TTT GAC GCT AAC TCC TGT GCG GTT CTG AGC GCT ATC GGC AAA 823
Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile Gly Lys
85 90 95

CAG TCT CCT GAC ATC AAC CAG GGC GTT GAC CGT GCC GAT CCG CTG GAA 871
Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro Leu Glu
100 105 110

CAG GGC GCG GGT GAC CAG GGT CTG ATG TTT GGC TAC GCA ACT AAT GAA 919
Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr Asn Glu
115 120 125 130

ACC GAC GTG CTG ATG CCA GCA CCT ATC ACC TAT GCA CAC CGT CTG GTA 967
Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg Leu Val
135 140 145

CAG CGT CAG GCT GAA GTG CGT AAA AAC GGC ACT CTG CCG TGG CTG CGC 1015
Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp Leu Arg
150 155 160

CCG GAC GCG AAA AGC CAG GTG ACT TTT CAG TAT GAC GAC GGC AAA ATC 1063
Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly Lys Ile
165 170 175

GTT GGT ATC GAT GCT GTC GTG CTT TCC ACT CAG CAC TCT GAA GAG ATC 1111
Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu Glu Ile
180 185 190

GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC AAG CCA 1159
Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile Lys Pro
195 200 205 210

ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC ATC AAC 1207
Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe Ile Asn
215 220 225

CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG 1255
Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu
230 235 240

ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC 1303
Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His
245 250 255

GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC 1351
Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser
260 265 270

GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC 1399
Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly
275 280 285 290

CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT 1447
Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala
295 300 305

GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT 1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro
310 315 320

TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA 1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro
325 330 335

TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA 1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu
340 345 350

ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA 1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys
355 360 365 370

ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC 1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
375 380 385

CGGGCCCCCC CTCGAGGTCG ACGGTATCGA TAAGCTTGAT ATCGAATTCC TGCAGCCCGG 1751

GGGATCCCAT GGTACGCGTC GAGGAGTACC ATG AAC GTT TTT AAT CCC GCG CAG 1805
Met Asn Val Phe Asn Pro Ala Gln
1 5

TTT CGC GCC CAG TTT CCC GCA CTA CAG GAT GCG GGC GTC TAT CTC GAC 1853
Phe Arg Ala Gln Phe Pro Ala Leu Gln Asp Ala Gly Val Tyr Leu Asp
10 15 20

AGC GCC GCG ACC GCG CTT AAA CCT GAA GCC GTG GTT GAA GCC ACC CAA 1901
Ser Ala Ala Thr Ala Leu Lys Pro Glu Ala Val Val Glu Ala Thr Gln
25 30 35 40

CAG TTT TAC AGT CTG AGC GCC GGA AAC GTC CAT CGC AGC CAG TTT GCC 1949
Gln Phe Tyr Ser Leu Ser Ala Gly Asn Val His Arg Ser Gln Phe Ala
45 50 55

GAA GCC CAA CGC CTG ACC GCG CGT TAT GAA GCT GCA CGA GAG AAA GTG 1997
Glu Ala Gln Arg Leu Thr Ala Arg Tyr Glu Ala Ala Arg Glu Lys Val
60 65 70

GCG CAA TTA CTG AAT GCA CCG GAT GAT AAA ACT ATC GTC TGG ACG CGC 2045
Ala Gln Leu Leu Asn Ala Pro Asp Asp Lys Thr Ile Val Trp Thr Arg
75 80 85

GGC ACC ACT GAA TCC ATC AAC ATG GTG GCA CAA TGC TAT GCG CGT CCG 2093
Gly Thr Thr Glu Ser Ile Asn Met Val Ala Gln Cys Tyr Ala Arg Pro
90 95 100

CGT CTG CAA CCG GGC GAT GAG ATT ATT GTC AGC GTG GCA GAA CAC CAC 2141
Arg Leu Gln Pro Gly Asp Glu Ile Ile Val Ser Val Ala Glu His His
105 110 115 120

GCC AAC CTC GTC CCC TGG CTG ATG GTC GCC CAA CAA ACT GGA GCC AAA 2189
Ala Asn Leu Val Pro Trp Leu Met Val Ala Gln Gln Thr Gly Ala Lys
125 130 135

GTG GTG AAA TTG CCG CTT AAT GCG CAG CGA CTG CCG GAT GTC GAT TTG 2237
Val Val Lys Leu Pro Leu Asn Ala Gln Arg Leu Pro Asp Val Asp Leu
140 145 150

TTG CCA GAA CTG ATT ACT CCC CGT AGT CGG ATT CTG GCG TTG GGT CAG 2285
Leu Pro Glu Leu Ile Thr Pro Arg Ser Arg Ile Leu Ala Leu Gly Gln
155 160 165

ATG TCG AAC GTT ACT GGC GGT TGC CCG GAT CTG GCG CGA GCG ATT ACC 2333
Met Ser Asn Val Thr Gly Gly Cys Pro Asp Leu Ala Arg Ala Ile Thr
170 175 180

TTT GCT CAT TCA GCC GGG ATG GTG GTG ATG GTT GAT GGT GCT CAG GGG 2381
Phe Ala His Ser Ala Gly Met Val Val Met Val Asp Gly Ala Gln Gly
185 190 195 200

GCA GTG CAT TTC CCC GCG GAT GTT CAG CAA CTG GAT ATT GAT TTC TAT 2429
Ala Val His Phe Pro Ala Asp Val Gln Gln Leu Asp Ile Asp Phe Tyr
205 210 215

GCT TTT TCA GGT CAC AAA CTG TAT GGC CCG ACA GGT ATC GGC GTG CTG 2477
Ala Phe Ser Gly His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Val Leu
220 225 230

TAT GGT AAA TCA GAA CTG CTG GAG GCG ATG TCG CCC TGG CTG GGC GGC 2525
Tyr Gly Lys Ser Glu Leu Leu Glu Ala Met Ser Pro Trp Leu Gly Gly
235 240 245

GGC AAA ATG GTT CAC GAA GTG AGT TTT GAC GGC TTC ACG ACT CAA TCT 2573
Gly Lys Met Val His Glu Val Ser Phe Asp Gly Phe Thr Thr Gln Ser
250 255 260

GCG CCG TGG AAA CTG GAA GCT GGA ACG CCA AAT GTC GCT GGT GTC ATA 2621
Ala Pro Trp Lys Leu Glu Ala Gly Thr Pro Asn Val Ala Gly Val Ile
265 270 275 280

GGA TTA AGC GCG GCG CTG GAA TGG CTG GCA GAT TAC GAT ATC AAC CAG 2669
Gly Leu Ser Ala Ala Leu Glu Trp Leu Ala Asp Tyr Asp Ile Asn Gln
285 290 295

GCC GAA AGC TGG AGC CGT AGC TTA GCA ACG CTG GCG GAA GAT GCG CTG 2717
Ala Glu Ser Trp Ser Arg Ser Leu Ala Thr Leu Ala Glu Asp Ala Leu
300 305 310

GCG AAA CGT CCC GGC TTT CGT TCA TTC CGC TGC CAG GAT TCC AGC CTG 2765
Ala Lys Arg Pro Gly Phe Arg Ser Phe Arg Cys Gln Asp Ser Ser Leu
315 320 325

CTG GCC TTT GAT TTT GCT GGC GTT CAT CAT AGC GAT ATG GTG ACG CTG 2813
Leu Ala Phe Asp Phe Ala Gly Val His His Ser Asp Met Val Thr Leu
330 335 340

CTG GCG GAG TAC GGT ATT GCC CTG CGG GCC GGG CAG CAT TGC GCT CAG 2861
Leu Ala Glu Tyr Gly Ile Ala Leu Arg Ala Gly Gln His Cys Ala Gln
345 350 355 360

CCG CTA CTG GCA GAA TTA GGC GTA ACC GGC ACA CTG CGC GCC TCT TTT 2909
Pro Leu Leu Ala Glu Leu Gly Val Thr Gly Thr Leu Arg Ala Ser Phe
365 370 375

GCG CCA TAT AAT ACA AAG AGT GAT GTG GAT GCG CTG GTG AAT GCC GTT 2957
Ala Pro Tyr Asn Thr Lys Ser Asp Val Asp Ala Leu Val Asn Ala Val
380 385 390

GAC CGC GCG CTG GAA TTA TTG GTG GAT TAAACGCGTG CTAGAGGCAT 3004
Asp Arg Ala Leu Glu Leu Leu Val Asp
395 400

CAAATAAAAC GAAAGGCTCA GTCGAAAGAC TGGGCCTTTC GTTTTATCTG TTGTTTGTCG 3064

GTGAACGCTC TCCTGAGTAG GACAAATCCG CCGCCCTAGA CCTAGGGGAT ATATTCCGCT 3124

TCCTCGCTCA CTGACTCGCT ACGCTCGGTC GTTCGACTGC GGCGAGCGGA AATGGCTTAC 3184

GAACGGGGCG GAGATTTCCT GGAAGATGCC AGGAAGATAC TTAACAGGGA AGTGAGAGGG 3244

CCGCGGCAAA GCCGTTTTTC CATAGGCTCC GCCCCCCTGA CAAGCATCAC GAAATCTGAC 3304

GCTCAAATCA GTGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG 3364

GCGGCTCCCT CGTGCGCTCT CCTGTTCCTG CCTTTCGGTT TACCGGTGTC ATTCCGCTGT 3424

TATGGCCGCG TTTGTCTCAT TCCACGCCTG ACACTCAGTT CCGGGTAGGC AGTTCGCTCC 3484

AAGCTGGACT GTATGCACGA ACCCCCCGTT CAGTCCGACC GCTGCGCCTT ATCCGGTAAC 3544

TATCGTCTTG AGTCCAACCC GGAAAGACAT GCAAAAGCAC CACTGGCAGC AGCCACTGGT 3604

AATTGATTTA GAGGAGTTAG TCTTGAAGTC ATGCGCCGGT TAAGGCTAAA CTGAAAGGAC 3664

AAGTTTTGGT GACTGCGCTC CTCCAAGCCA GTTACCTCGG TTCAAAGAGT TGGTAGCTCA 3724

GAGAACCTTC GAAAAACCGC CCTGCAAGGC GGTTTTTTCG TTTTCAGAGC AAGAGATTAC 3784

GCGCAGACCA AAACGATCTC AAGAAGATCA TCTTATTAAT CAGATAAAAT ATTTCTAGAT 3844

TTCAGTGCAA TTTATCTCTT CAAATGTAGC ACCTGAAGTC AGCCCCATAC GATATAAGTT 3904

GTTACTAGTG CTTGGATTCT CACCAATAAA AAACGCCCGG CGGCAACCGA GCGTTCTGAA 3964

CAAATCCAGA TGGAGTTCTG AGGTCATTAC TGGATCTATC AACAGGAGTC CAAGCGAGCT 4024

CTCGAACCCC AGAGTCCCGC TCAGAAGAAC TCGTCAAGAA GGCGATAGAA GGCGATGCGC 4084

TGCGAATCGG GAGCGGCGAT ACCGTAAAGC ACGAGGAAGC GGTCAGCCCA TTCGCCGCCA 4144

AGCTCTTCAG CAATATCACG GGTAGCCAAC GCTATGTCCT GATAGCGGTC CGCCACACCC 4204

AGCCGGCCAC AGTCGATGAA TCCAGAAAAG CGGCCATTTT CCACCATGAT ATTCGGCAAG 4264

CAGGCATCGC CATGGGTCAC GACGAGATCC TCGCCGTCGG GCATGCGCGC CTTGAGCCTG 4324

GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA 4384

AGACCGGCTT CCATCCGAGT ACGTGCTCGC TCGATGCGAT GTTTCGCTTG GTGGTCGAAT 4444

GGGCAGGTAG CCGGATCAAG CGTATGCAGC CGCCGCATTG CATCAGCCAT GATGGATACT 4504

TTCTCGGCAG GAGCAAGGTG AGATGACAGG AGATCCTGCC CCGGCACTTC GCCCAATAGC 4564

AGCCAGTCCC TTCCCGCTTC AGTGACAACG TCGAGCACAG CTGCGCAAGG AACGCCCGTC 4624

GTGGCCAGCC ACGATAGCCG CGCTGCCTCG TCCTGCAGTT CATTCAGGGC ACCGGACAGG 4684

TCGGTCTTGA CAAAAAGAAC CGGGCGCCCC TGCGCTGACA GCCGGAACAC GGCGGCATCA 4744

GAGCAGCCGA TTGTCTGTTG TGCCCAGTCA TAGCCGAATA GCCTCTCCAC CCAAGCGGCC 4804

GGAGAACCTG CGTGCAATCC ATCTTGTTCA ATCATGCGAA ACGATCCTCA TCCTGTCTCT 4864

TGATCAGATC TTGATCCCCT GCGCCATCAG ATCCTTGGCG GCAAGAAAGC CATCCAGTTT 4924

ACTTTGCAGG GCTTCCCAAC CTTACCAGAG GGCGCCCCAG CTGGCAATTC C 4975

(2) INFORMATION FOR SEQ ID No: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 14:

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
1 5 10 15

Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
20 25 30

Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
35 40 45

Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
50 55 60

Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
65 70 75 80

Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
85 90 95

Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
100 105 110

Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
115 120 125

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
130 135 140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
165 170 175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
180 185 190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

(2) INFORMATION FOR SEQ ID No: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 401 Amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 15:

```
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
 1               5                  10                  15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
            20                  25                  30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
        35                  40                  45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
    50                  55                  60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65                  70                  75                  80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
                85                  90                  95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
            100                 105                 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
        115                 120                 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
    130                 135                 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145                 150                 155                 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
                165                 170                 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
            180                 185                 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
        195                 200                 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
    210                 215                 220
```

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

Asp

# *Declaration, Power of Attorney*



*0050/048792*

We (I), the undersigned inventor(s), hereby declare(s) that:

My residence, post office address and citizenship are as stated below next to my name,

We (I) believe that we are (I am) the original, first, and joint (sole) inventor(s) of the subject matter which is claimed and for which a patent is sought on the invention entitled

Process for preparing biotin

the specification of which

[ ] is attached hereto.

[ ] was filed on ______________________ as

Application Serial No. ______________________

and amended on ______________________.

[x] was filed as PCT international application

Number PCT/EP 99/01052

on 17/02/1999

and was amended under PCT Article 19

on ______________________ (if applicable).

We (I) hereby state that we (I) have reviewed and understand the contents of the above–identified specification, including the claims, as amended by any amendment referred to above.

We (I) acknowledge the duty to disclose information known to be material to the patentability of this application as defined in Section 1.56 of Title 37 Code of Federal Regulations.

We (I) hereby claim foreign priority benefits under 35 U.S.C. § 119(a)–(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed. Prior Foreign Application(s)

| **Application No.** | **Country** | **Day/Month/Year** | **Priority Claimed** |
|---|---|---|---|
| *19806872.7* | *Germany* | *19 February 1998* | [x] Yes [ ] No |

pdtuskpct001 - 25

We (I) hereby claim the benefit under Title 35, United States Codes, § 119(e) of any United States provisional application(s) listed below.

| | |
|---|---|
| (Application Number) | (Filing Date) |
| (Application Number) | (Filing Date) |

We (I) hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s), or § 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR § 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application.

| Application Serial No. | Filing Date | Status (pending, patented, abandoned) |
|---|---|---|
| | | |
| | | |
| | | |
| | | |

2 And we (I) hereby appoint **Messrs. HERBERT. B. KEIL**, Registration Number 18,967; and **RUSSEL E. WEINKAUF**, Registration Number 18,495; the address of both being Messrs. Keil & Weinkauf, 1101 Connecticut Ave., N.W., Washington, D.C. 20036 (telephone 202–659–0100), our attorneys, with full power of substitution and revocation, to prosecute this application, to make alterations and amendments therein, to sign the drawings, to receive the patent, and to transact all business in the Patent Office connected therewith.

We (I) declare that all statements made herein of our (my) own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

1-00 *Hartwig Schröder*
NAME OF INVENTOR

Hartwig Schröder
Signature of Inventor

Date
04/03/1999

*Goethestr.5*
*69226 Nußloch* DEX
*Germany*
*Citizen of: Germany*
Post Office Address: same as residence

09529140

pduskpet001 - 25